



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CLEBER ZANCHETTIN

“SISTEMA NEURAL HÍBRIDO PARA RECONHECIMENTO
DE PADRÕES EM UM NARIZ ARTIFICIAL”

*ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO
EM CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE
INFORMÁTICA DA UNIVERSIDADE FEDERAL DE
PERNAMBUCO COMO REQUISITO PARCIAL PARA
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA
COMPUTAÇÃO.*

ORIENTADOR(A): PROFA. DRA. TERESA BERNARDA LUDERMIR

RECIFE, MARÇO/2004

CLEBER ZANCHETTIN

**SISTEMA NEURAL HÍBRIDO PARA RECONHECIMENTO DE
PADRÕES EM UM NARIZ ARTIFICIAL**

Dissertação apresentada ao Curso de Mestrado em
Ciência da Computação como requisito parcial à
obtenção do grau de Mestre em Ciência da
Computação

Universidade Federal de Pernambuco

Orientadora: Profa. Dra. Teresa Bernarda Ludermir

RECIFE, 2004

DEDICATÓRIAS

para minha família.

AGRADECIMENTOS

Agradeço a todos que auxiliaram, direta ou indiretamente, no desenvolvimento desse trabalho, tornando-o uma realidade ao contrário de uma pretensão.

Especialmente a professora Teresa Bernarda Ludermir, por ter acreditado e confiado em mim desde o momento em que nos conhecemos.

Aos meus irmãos Joel da Silva e Tarcisio Pinto Camara pelas infindáveis discussões acadêmicas, pessoais e culinárias.

Como disse Max Planck: “Para os crentes Deus está no princípio das coisas. Para os cientistas no final de toda reflexão”.

Todos vocês são muito importantes para mim.

SISTEMA NEURAL HÍBRIDO PARA RECONHECIMENTO DE PADRÕES EM UM NARIZ ARTIFICIAL¹

Autor: Cleber Zanchettin

Orientadora: Dra. Teresa Bernarda Ludermir

RESUMO

Esta dissertação investiga a utilização de Sistemas Híbridos Inteligentes no sistema de reconhecimento de padrões de um nariz artificial. O trabalho envolve cinco partes principais: (1) avaliação da base de dados de odores a partir de uma técnica estatística multivariada; (2) validação das Redes Neurais com Atrasos no Tempo no reconhecimento de odores; (3) avaliação da Transformada de Wavelets como método de pré-processamento de sinais de odores em abordagens conexionistas; (4) avaliação de abordagens híbridas inteligentes para o reconhecimento de odores em narizes artificiais; e (5) o estudo de caso.

Duas arquiteturas híbridas inteligentes foram investigadas em detalhes: a rede neuro-difusa Feature-weighted Detector, que permite a classificação de padrões, seleção dos atributos mais importantes e extração de regras explicativas da rede; e a rede neuro-difusa Evolving Fuzzy Neural Networks, que possibilita o aprendizado *on-line* e incremental, além da inserção, extração e agregação de conhecimento em sua arquitetura evolutiva.

Foram analisados sinais gerados por um nariz artificial, composto por um conjunto de oito sensores de polímeros condutores, exposto a gases derivados do petróleo. A utilização da Transformada de Wavelet melhorou consideravelmente o desempenho dos classificadores conexionistas. Nos experimentos realizados, as Redes Neurais com Atrasos no Tempo obtiveram um erro médio de classificação de 0.75%, enquanto que as Redes Perceptron Multi-Camadas obtiveram um erro médio de 11.5%. Dentre as abordagens híbridas investigadas, a rede Feature-weighted Detector, obteve um erro médio de classificação de 20.72% e a rede Evolving Fuzzy Neural Networks obteve um erro médio de 0.88% na classificação dos odores.

Palavras-chave: Sistemas Híbridos Inteligentes, Reconhecimento de Padrões, Narizes Artificiais, Redes Neurais Artificiais.

¹ Dissertação de Mestrado em Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, (138p.), Março de 2004.

HYBRID NEURAL SYSTEM FOR PATTERN RECOGNITION IN AN ARTIFICIAL NOSE²

Author: Cleber Zanchettin

Adviser: Dra. Teresa Bernarda Ludermir

ABSTRACT

This dissertation investigates the use of Hybrid Intelligent Systems in the pattern recognition system of an artificial nose. The work involves five main parts: (1) an evaluation of the odors database by a multivariate statistics technique; (2) a validation of the Time Delay Neural Networks in the odors recognition; (3) an evaluation of the Wavelet Transform as preprocessing method of odors signals in the connectionist approaches; (4) an evaluation of hybrid intelligent systems for the odor recognition in artificial noses; and (5) a case study.

Two intelligent hybrid architectures had been investigated: the neuro-fuzzy Feature-weighted Detector network, which allows the pattern classification, feature selection and rule extration of the network; and the neuro-fuzzy Evolving Fuzzy Neural Networks, that enables on-line and incremental learning, insertion, extration and aggregation of knowledge in its evolutive architecture.

The signals generated by an artificial nose, composed by an array of eight conducting polymer sensors, exposed to the gases derived from the petroliferous industry were analyzed. The use of the Wavelet Transform improved the performance of the connectionist classifiers. In the experiments, the Time Delay Neural Networks obtained a mean classification error of 0.75%, while the Multi-Layer Perceptron obtained an error of 11.5%. In the investigated hybrid systems, the Feature-weighted Detector obtained a mean classification error of 20.72%. The Evolving Neural Fuzzy Networks obtained a mean classification error of 0.88% in the odors classification.

Keywords: Hybrid Intelligent Systems, Pattern Recognition, Artificial Noses, Artificial Neural Networks.

² Master of Science dissertation in Computer Science, Informatics Center, Federal University of Pernambuco, Recife, PE, (138p.), March, 2004.

SUMÁRIO

<i>Introdução</i>	<i>1</i>
1.1 Motivação	1
1.2 Objetivos da dissertação	4
1.3 Descrição do trabalho	5
1.4 Organização da dissertação	7
<i>Narizes Artificiais</i>	<i>9</i>
2.1 Introdução	9
2.2 Sistema sensor	11
2.2.1 Sensores de absorção	11
2.2.2 Semicondutores quimicamente sensíveis	12
2.3 Pré-processamento	13
2.4 Sistemas de reconhecimento de padrões para narizes artificiais	15
2.4.1 Análise de Função Discriminante	16
2.4.2 Análise de Agrupamentos	16
2.4.3 Análise de Componentes Principais	17
2.4.4 Redes Neurais Artificiais	17
2.5 Sumário do capítulo	19
<i>Descrição do Problema e Análise dos Dados</i>	<i>20</i>
3.1 Aquisição dos dados	20
3.2 Análise dos dados	23
3.3 Sumário do Capítulo	27
<i>Sistemas de Reconhecimento de Padrões para Narizes Artificiais</i>	<i>29</i>
4.1 Abordagens conexionistas para reconhecimento de padrões	29
4.1.1 Redes Neurais Perceptron Multi-Camadas	30
4.1.2 Redes Neurais com Atrasos no Tempo	32

4.2 Abordagens neurais híbridas para o reconhecimento de padrões	35
4.2.1 Rede Feature-weighted Detector	36
4.2.2 Rede Evolving Fuzzy Neural Network	42
4.3 Sumário do capítulo	53
Métodos	55
5.1 Pré-processamento dos dados	55
5.1.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste	55
5.1.2 Normalização	57
5.1.3 Transformada de Wavelets	58
5.2 Aspectos observados	68
5.3 Métodos de comparação entre as abordagens	68
5.4 Sumário do capítulo	70
Experimentos	71
6.1 Experimentos com a rede Perceptron Multi-Camadas	71
6.1.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste	71
6.1.2 Pré-processamento dos dados	72
6.1.3 Topologia utilizada	72
6.1.4 Metodologia de treinamento	73
6.1.5 Resultados obtidos	74
6.2 Experimentos com a Rede Neural com Atrasos no Tempo	79
6.2.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste	79
6.2.2 Pré-processamento	80
6.2.3 Topologia utilizada	81
6.2.4 Metodologia de treinamento	81
6.2.5 Resultados obtidos	82
6.3 Experimentos com a rede Feature-weighted Detector	86
6.3.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste	86
6.3.2 Pré-processamento	87
6.3.3 Topologia utilizada	88
6.3.4 Metodologia de treinamento	88
6.3.5 Resultados obtidos	88
6.4 Experimentos com a rede Evolving Fuzzy Neural Networks	94
6.4.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste	95
6.4.2 Pré-processamento	95
6.4.3 Topologia utilizada	96
6.4.4 Metodologia de treinamento	96
6.4.5 Resultados obtidos	97
6.5 Comparação entre as abordagens	104
6.5.1 Rede Perceptron Multi-Camadas	106
6.5.2 Rede Neural com Atrasos no Tempo	107
6.5.3 Rede Feature-weighted Detector	108
6.5.4 Rede Evolving Fuzzy Neural Networks	109
6.6 Sumário do Capítulo	111
Conclusão	112
7.1 Considerações finais	112

7.2 Contribuições deste trabalho	115
7.3 Proposta para trabalhos futuros	116
<i>Apêndice A</i>	<i>118</i>
<i>Apêndice B</i>	<i>122</i>
<i>Referências Bibliográficas</i>	<i>130</i>

LISTA DE FIGURAS

<i>Figura 2.1 - Estrutura básica de um nariz artificial</i>	11
<i>Figura 3.1- Curva temporal formada pelos valores de resistência do gás Butano</i>	23
<i>Figura 3.2 - Variância explicada em cada uma das componentes principais</i>	25
<i>Figura 3.3 - Projeção dos dados em três componentes principais</i>	26
<i>Figura 3.4 - Correlação entre os sensores</i>	27
<i>Figura 4.1 - Exemplo de uma rede Perceptron Multi-Camadas</i>	31
<i>Figura 4.2 - Método de janelamento para processamento temporal</i>	33
<i>Figura 4.3 - Rede Time Delay Neural Network para processamento temporal</i>	34
<i>Figura 4.4 - Modelo da rede Feature-weighted Detector</i>	37
<i>Figura 4.5 - Comportamento estrutural e de interconexão do neurônio i</i>	38
<i>Figura 4.6 - Arquitetura da rede Evolving Fuzzy Neural Network (adaptado de Kasabov (2001))</i>	44
<i>Figura 4.7 - Associação entre hiper-esferas de entrada e saída</i>	45
<i>Figura 4.8 - Processo de agregação e otimização da arquitetura (adaptado de Kasabov (2001))</i>	48
<i>Figura 4.9 - Agregação de nodos de regras em uma rede Evolving Fuzzy Neural Network</i>	49
<i>Figura 5.1 - Ilustração de como são calculados os coeficientes da Transformada de Wavelet</i>	61
<i>Figura 5.2 - Árvore de decomposição tradicional da transformada wavelet com seis níveis de decomposição (adaptado de [Misiti, 1996]).</i>	62
<i>Figura 5.3 - Decomposição do sinal, s, nas componentes terminais da árvore da Figura 5.2</i>	63
<i>Figura 5.4 - Árvore de decomposição utilizada para filtrar o ruído e comprimir o sinal.</i>	66
<i>Figura 5.5 - Decomposição até o nível 3: $s = a_3 + d_3 + d_2 + d_1$.</i>	67
<i>Figura 5.6 - Transformada Wavelet: Coeficientes, Sinais e Aproximações.</i>	67
<i>Figura 6.1 - Projeção das saídas obtidas e desejadas nos experimentos realizados com a Partição 1</i>	98
<i>Figura 6.2 - Aplicação da técnica de poda nos experimentos com a Partição 2</i>	100
<i>Figura 6.3 - Visualização dos padrões e seus respectivos nodos de regras, espaço original</i>	101
<i>Figura 6.4 - Visualização dos padrões e seus respectivos nodos de regras, espaço PCA</i>	101
<i>Figura A.1 - Exemplo da extração dos componentes principais [Daniels, 2003]</i>	119
<i>Figura B.1 - Teste de hipótese bilateral</i>	126
<i>Figura B.2 - Teste de hipótese unilateral à esquerda</i>	126
<i>Figura B.3 - Teste de hipótese unilateral à direita</i>	127
<i>Figura B.4 - Distribuição t de Student</i>	128

LISTA DE TABELAS

<i>Tabela 3.1 - Descrição dos sensores utilizados no nariz artificial</i>	21
<i>Tabela 3.2 - Exemplo de aquisição dos dados para o gás Butano</i>	22
<i>Tabela 3.3 - Variância descrita pelas Componentes Principais</i>	24
<i>Tabela 3.4 - Peso de cada variável nas Componentes Principais</i>	27
<i>Tabela 6.1 - MLP - Média e desvio-padrão para Partição 1</i>	75
<i>Tabela 6.2 - MLP - Média e desvio-padrão para Partição 2</i>	76
<i>Tabela 6.3 - MLP - Média e desvio-padrão para partição 3</i>	76
<i>Tabela 6.4 - MLP com Wavelet - Média e desvio-padrão para a Partição 1</i>	77
<i>Tabela 6.5 - MLP com Wavelet - Média e desvio-padrão para a Partição 2</i>	77
<i>Tabela 6.6 - MLP com Wavelet - Média e desvio-padrão para a Partição 3</i>	77
<i>Tabela 6.7 - Estatística do teste emparelhado t de Student</i>	79
<i>Tabela 6.8 - TDNN - Média e desvio-padrão para a Partição 1</i>	82
<i>Tabela 6.9 - TDNN - Média e desvio-padrão para a Partição 2</i>	83
<i>Tabela 6.10 - TDNN - Média e desvio-padrão para a Partição 3</i>	83
<i>Tabela 6.11 - TDNN com Wavelet - Média e desvio-padrão para a Partição 1</i>	84
<i>Tabela 6.12 - TDNN com Wavelet - Média e desvio-padrão para a Partição 2</i>	84
<i>Tabela 6.13 - TDNN com Wavelet - Média e desvio-padrão para a Partição 3</i>	85
<i>Tabela 6.14 - Estatística do teste emparelhado t de Student</i>	85
<i>Tabela 6.15 - Média e desvio-padrão dos resultados obtidos com a rede Feature-weighted Detector</i>	88
<i>Tabela 6.16 - Médias dos valores finais para as conexões de memória</i>	89
<i>Tabela 6.17 - Médias dos valores finais para as conexões de pesos</i>	90
<i>Tabela 6.18 - Resultados da rede Feature-weighted Detector na classificação de sinais de odores</i>	91
<i>Tabela 6.19 - Comparação dos resultados da seleção de atributos</i>	92
<i>Tabela 6.20 - Estatística do teste emparelhado t de Student</i>	93
<i>Tabela 6.21 - Resultados da Evolving Fuzzy Neural Network – simulação 1</i>	97
<i>Tabela 6.22 - Resultados da Evolving Fuzzy Neural Network – simulação 2</i>	99
<i>Tabela 6.23 - Resultados da Evolving Fuzzy Neural Network – simulação 3</i>	100
<i>Tabela 6.24 - Resultados da Evolving Fuzzy Neural Network – simulação 4</i>	103
<i>Tabela 6.25 - Melhores resultados obtidos pelos modelos investigados</i>	105
<i>Tabela 6.26 - Comparação entre os classificadores</i>	106
<i>Tabela B.1 - Tipos de teste de hipótese</i>	123
<i>Tabela B.2 - Tipos de erros</i>	125

1

Introdução

Este capítulo apresenta a motivação desta dissertação, analisando aspectos inerentes ao processo de reconhecimento de odores em narizes artificiais. A seção 1.1 expõe brevemente o problema de reconhecimento de odores, o qual será aprofundado na dissertação. A seção 1.2 demarca os objetivos deste trabalho. A seção 1.3 contém uma sucinta descrição do trabalho realizado na dissertação. E a seção 1.4 fornece uma visão dos capítulos da dissertação.

1.1 Motivação

A busca por dispositivos artificiais que apresentem algum tipo de comportamento inteligente, similar ao exibido por muitos organismos biológicos, sempre fascinou a ciência. Na pesquisa por sistemas que possam exibir características biológicas, uma das mais intensas batalhas que o homem tem travado em seu progresso científico e tecnológico é tentar reproduzir funções dos órgãos dos sentidos humanos; com o objetivo de criar dispositivos eletrônicos aptos a substituir, simular, aumentar o poder ou estender os sentidos humanos além de seus limites naturais.

A partir de pesquisas realizadas com os órgãos sensoriais humanos foi possível construir aparelhos que, embora com as devidas limitações, reproduzam o comportamento de alguns sentidos como o ouvido (o microfone), a visão (a câmera) e o tato (sensores de pressão e calor). Além do desenvolvimento de outros dispositivos que

desempenham funções diferentes das dos órgãos sensoriais, mas que estão relacionadas com eles, como a televisão e o alto-falante.

No entanto dois órgãos dos sentidos, o paladar e o olfato ficaram durante muito tempo esquecidos, perante o progresso científico obtido na pesquisa dos demais órgãos sensoriais. Esta omissão, em especial do sistema olfativo, se deve justamente por muitos dos aspectos inerentes ao processo de reconhecimento de odores em sistemas biológicos ainda não estarem bem esclarecidos pela ciência [Dos Santos, 1999]. Este fato motivou o estudo dos processos envolvidos na olfação humana; e mesmo sem a completa compreensão da fisiologia olfativa biológica, impulsionaram o desenvolvimento de vários modelos de sistemas olfativos artificiais baseados em evidências biológicas [Li e Hopfield, 1989; Li e Hertz, 2000].

Os modelos olfativos artificiais, inspirados em sistemas olfativos biológicos, originaram dispositivos eletrônicos hoje conhecidos como *Narizes Artificiais*³. Os narizes artificiais são sistemas capazes de detectar e classificar odores, vapores e gases automaticamente; atualmente são constituídos por dois módulos fundamentais: o *Sistema Sensor*, responsável por efetuar a medição do composto odorante no ambiente [Yamazaki, 2001; Santos, 2000 e Keller et al., 1995; Dos Santos, 1999; Gardner et al., 1994]; e o *Sistema de Reconhecimento de Padrões*, responsável pelo reconhecimento e classificação dos sinais de odores detectados pelo nariz artificial [Gardner e Bartlett, 1992 e Gardner, 1991; Yamazaki, 2001; Santos, 2000].

Os narizes artificiais têm sido utilizados com grande sucesso em diversas aplicações [Keller et al., 1995], e têm despertado especial interesse da indústria petrolífera, onde podem ser utilizados como dispositivos de segurança na detecção e reconhecimento de gases tóxicos. Em companhias petrolíferas, os narizes artificiais podem atuar na prevenção de vazamentos em dutos de transporte e em plataformas de extração e refino de derivados de petróleo. Posicionados em locais estratégicos, os narizes artificiais podem detectar e identificar vazamentos químicos em estágios iniciais, antes que a concentração de gases tóxicos alcance níveis perceptíveis e perigosos para os seres humanos e para a natureza.

³ Estes dispositivos também são referenciados na literatura como Narizes Eletrônicos.

A detecção e identificação de substâncias derivadas do petróleo são tarefas de extrema importância, tanto no campo de extração e produção, quanto nas plantas de refino e processamento de derivados do petróleo. Razões econômicas, de segurança e de controle ambiental, fazem com que cada vez mais sejam despendidos esforços no sentido de desenvolver tecnologias e equipamentos para verificar a presença de gases tóxicos em ambientes controlados.

Tendo grande importância nos resultados produzidos por um nariz artificial, o sistema de reconhecimento de padrões tem recebido crescente atenção em seu desenvolvimento. Igualmente importante para diversas áreas, o reconhecimento de padrões é um campo em franca expansão no domínio da Inteligência Computacional. Dentre as abordagens investigadas objetivando o reconhecimento de padrões, resultados promissores estão sendo encontrados no desenvolvimento de *Sistemas Híbridos Inteligentes* [Prudêncio, 2002; Abraham e Nath, 2000; Braga et al., 2000 ; McGarry et al., 1999; Goonatilake e Khebbal, 1995; Medsker, 1995]. Os sistemas híbridos inteligentes são modelos que resultam da combinação de duas ou mais técnicas computacionais distintas, visando combinar vantagens para suprir deficiências individuais, na resolução de determinado problema.

Na pesquisa por sistemas que possam compartilhar ou combinar características, diversos pesquisadores têm constatado que técnicas de aprendizagem de dados são, em muitos aspectos, complementares. Assim, muitos esforços têm sido realizados no sentido de investigar possíveis formas de cooperação entre métodos inteligentes ou mesmo entre métodos de inteligência computacional e métodos estatísticos ou de engenharia [Abraham e Nath, 2000; Zanchettin e Ludermir, 2003a; Kasabov, 1988; Jang, 1993; Kandel et al., 1998; Nauck e Kursel, 1999].

Neste trabalho, procuramos investigar técnicas híbridas que possam contribuir para o aprimoramento do sistema de reconhecimento de padrões do protótipo de um nariz artificial [Santos, 2000], a ser utilizado na detecção de gases provenientes da indústria petrolífera. São abordadas redes neurais artificiais já consagradas na literatura no reconhecimento de odores, como redes neurais *Perceptron Multi-Camadas* [Rumelhart et al., 1986] e *Redes Neurais com Atraso no Tempo* [Lang e Hilton, 1988];

além de novas abordagens neurais híbridas como: a rede neuro-difusa *Feature-weighted Detector* [Li et al., 2002], capaz de efetuar a classificação de padrões, seleção dos atributos mais importantes e extração de regras difusas da rede; e a rede neural evolutiva *Evolving Fuzzy Neural Network* [Kasabov, 1999], que opera de um modo *on-line*; aprende incrementalmente e manipula o conhecimento adquirido do ambiente a partir de regras difusas, realizando agregação, inserção e extração destas regras na arquitetura da rede.

1.2 Objetivos da dissertação

Este trabalho estuda sistemas de reconhecimento de padrões para narizes artificiais, bem como as diversas abordagens já apresentadas na literatura para este fim, apresentado suas vantagens e desvantagens.

Dentre os objetivos específicos deste trabalho, está a apresentação de uma análise estatística dos dados do problema, descrito no Capítulo 3. Esta investigação é realizada a partir de uma técnica estatística de análise de multivariáveis, que tem por objetivo a interpretação, visualização e redução da dimensionalidade dos dados.

Objetiva-se ainda contribuir para a validação dos resultados apresentados em trabalhos anteriores [Yamazaki et al., 2001 e Yamazaki e Ludermir, 2001], a partir de uma nova base de dados. A abordagem proposta por Yamazaki (2001) considera características temporais, dos sinais gerados pelos sensores do nariz artificial, na classificação dos padrões de odores.

A fim de comparar os resultados das abordagens neurais híbridas investigadas, com outros modelos utilizados no reconhecimento de sinais de odores, serão também apresentados resultados da aplicação das redes neurais *Perceptron Multi-Camadas* no reconhecimento dos sinais de odores da referida base de dados.

Nesta dissertação se propõe um Sistema de Reconhecimento de Padrões baseado em abordagens neurais híbridas, para a classificação dos sinais de odores provenientes dos sensores do nariz artificial.

Vale ressaltar que o principal objetivo dessa pesquisa, é a busca por sistemas que sejam eficientes, robustos, fáceis de operar, e capazes de fornecer soluções de qualidade para problemas complexos.

1.3 Descrição do trabalho

Este trabalho consistiu das seguintes etapas: (1) análise dos padrões de odores provenientes de um protótipo de nariz artificial; (2) seleção e implementação de modelos neurais híbridos para construção de um sistema de reconhecimento de padrões para um nariz artificial; e (3) estudo de caso, com a aplicação dos modelos investigados na classificação de gases provenientes da indústria petrolífera.

A fase de análise dos padrões de odores compreendeu a uma descrição da base de dados dos gases derivados do petróleo. Esta base foi examinada com uma técnica de análise multivariada, a qual tem por objetivo a interpretação, visualização e redução da dimensionalidade dos dados. Esta técnica, conhecida como *Análise de Componentes Principais* [Jackson, 1991], foi utilizada como método de visualização e interpretação dos sinais de odores, com o objetivo de verificar características e particularidades dos padrões que estão sendo investigados.

A seleção dos modelos neurais híbridos, para o sistema de reconhecimento de padrões do nariz artificial, englobou o estudo das técnicas utilizadas na literatura para classificação de odores. Verificou-se que a maioria das soluções conexionistas fazem uso da rede *Perceptron Multi-Camadas* estática, sem laços de realimentação, empregando o algoritmo Backpropagation [Rumelhart et al., 1986] para o aprendizado da rede.

Outro modelo conexionista que se destaca na literatura é a *Rede Neural com Atrasos no Tempo*, esta abordagem utiliza características temporais dos sinais gerados pelos sensores do nariz artificial, considerando a variação destes valores ao longo do intervalo de tempo em que foram feitas as aquisições dos dados. Resultados promissores foram obtidos com esta arquitetura no reconhecimento de sinais de odores. O que torna

esta técnica necessária, juntamente com a rede *Perceptron Multi-Camadas*, na comparação e avaliação das demais técnicas propostas para a classificação de odores nesta dissertação.

A fim de melhorar o desempenho de classificação das técnicas acima referenciadas, no reconhecimento dos sinais de odores, foram realizados testes com técnicas de pré-processamento de dados. As técnicas utilizadas foram: a *Normalização* [Andrade, 2002], uma técnica de transformação de dados; e a *Transformada de Wavelet* [Morlet, 1981], uma ferramenta de processamento de sinais digitais que foi utilizada como filtro para a eliminação do ruído e compressão dos sinais de odores.

Dentre as abordagens neurais híbridas verificadas, duas arquiteturas foram amplamente investigadas. A rede neuro-difusa *Feature-weighted Detector*, que consiste de uma rede com conexões de memória para classificação de padrões e conexões de pesos para seleção de atributos. Além da capacidade de selecionar atributos e classificar padrões a rede *Feature-weighted Detector* possibilita a extração do conhecimento, adquirido durante a fase de aprendizado da rede, na forma de um conjunto de regras interpretáveis do tipo *Se-Então*.

A outra arquitetura híbrida investigada foi a rede neural evolutiva *Evolving Fuzzy Neural Network*, que possui características como, aprendizado rápido utilizando regras difusas para seu crescimento, poda e agregação. Os nodos de processamento da rede *Evolving Fuzzy Neural Network* e as conexões entre eles são criados, eliminados e agregados dinamicamente durante a operação da rede. A partir destas funcionalidades esta arquitetura pode também inserir classes dinamicamente em sua topologia e se adaptar a novos dados, de uma maneira *on-line*.

Todos os métodos investigados foram implementados e testados com o programa de simulação *Matlab*⁴, utilizando o *toolbox de redes neurais* [Misiti, 1996] como base de implementação da rede *Perceptron Multi-Camadas*. Vale ressaltar que as demais abordagens pesquisadas foram implementadas a partir de especificações obtidas na revisão bibliográfica.

⁴ MATLAB, the Language of Technical Computing. Marca registrada da empresa The MathWorks, Inc.

No estudo de caso, foram tratados os padrões de odores, obtidos a partir da exposição do protótipo de um nariz artificial [Santos, 2000], frente a compostos fornecidos pela companhia petrolífera Petrobras⁵. Os gases analisados foram: Metano, Etano, Propano, Butano e Monóxido de Carbono.

1.4 Organização da dissertação

O Capítulo 2 apresenta uma introdução aos narizes artificiais. São descritas as partes constituintes de um nariz artificial: Sistema Sensor e Sistema de Reconhecimento de Padrões. A partir desta descrição são destacadas, em linhas gerais, as principais abordagens propostas na literatura para sistemas de reconhecimento de padrões em narizes artificiais.

No Capítulo 3 é realizada uma análise dos dados de odores utilizados nesta dissertação, a partir de uma técnica estatística multivariada. Esta técnica, a Análise de Componentes Principais, é aplicada sobre a base de dados de odores. Os resultados da análise são apresentados e investigados em detalhes, fornecendo um conhecimento prévio sobre o problema abordado.

Os modelos para o reconhecimento de padrões, investigados nesta dissertação, são introduzidas no Capítulo 4. Eles compreendem às técnicas conexionistas como: *Perceptron Multi-Camadas* e *Redes com Atrasos no Tempo*; e abordagens neurais híbridas como *Feature-weighted Detector* e *Evolving Fuzzy Neural Networks*. O funcionamento de cada uma destas redes é detalhadamente descrito, de forma que uma visão geral sobre cada uma das abordagens seja obtida.

O Capítulo 5 descreve a metodologia utilizada nos experimentos com as técnicas investigadas. Inicialmente, uma descrição sobre a metodologia dos experimentos é fornecida. Também são descritas as técnicas de pré-processamento de dados, utilizadas nos experimentos, que são a *Normalização* e a *Transformada de Wavelet*. Posterior a

⁵ Petróleo Brasileiro S.A., Brasil.

descrição das técnicas de pré-processamento são apresentadas considerações sobre os métodos de avaliação e teste de hipótese utilizados.

O Capítulo 6 descreve os experimentos realizados com cada uma das técnicas investigadas no reconhecimento dos padrões de odores. São expostos os experimentos e resultados de cada uma das abordagens.

As considerações finais sobre o trabalho apresentado nesta dissertação são tecidas no Capítulo 7. Além das considerações finais, este capítulo apresenta uma descrição das contribuições desta dissertação e algumas propostas para trabalhos futuros.

2

Narizes Artificiais

Este capítulo apresenta um breve histórico sobre o desenvolvimento de narizes artificiais, são analisadas suas partes constituintes, bem como as diversas abordagens existentes para cada componente. Na seção 2.1 um resumo sobre o desenvolvimento de narizes artificiais é apresentado. A seção 2.2 descreve o funcionamento do Sistema Sensor do nariz artificial. Na seção 2.3 são descritas as técnicas de pré-processamento utilizadas no tratamento de sinais de odores. Na seção 2.4 são descritas várias abordagens para Sistemas de Reconhecimento de Padrões em narizes artificiais. A seção 2.5 é um sumário do capítulo.

2.1 Introdução

Apesar da pesquisa e desenvolvimento de narizes artificiais serem inspirados no sistema olfativo biológico, sua implementação, não se aproxima nem da complexidade ou de muitas das funcionalidades de seu equivalente biológico. No sistema olfativo biológico, o processo de detecção de odores se inicia com a codificação do estímulo no epitélio olfativo. As moléculas de odor penetram na cavidade nasal e são difundidas na camada de muco para interagirem com os neurônios receptores. Esta interação gera um padrão de atividade que é transmitido para o bulbo olfativo, onde ocorrem as primeiras interações sinápticas. O padrão recebido é processado e transformado em um novo sinal, o qual é enviado para o córtex olfativo. No córtex olfativo, ocorrem associações que são

fundamentais para a discriminação do estímulo odorante. Em seguida, os sinais processados são enviados para outras partes do cérebro, como o tálamo e o hipotálamo, onde influenciam a consciência e o comportamento do indivíduo [Yamazaki, 2001; Li e Hertz, 2000 e Li e Hopfield, 1989].

Atualmente, o funcionamento dos narizes artificiais ocorre da seguinte maneira: o processo de reconhecimento e classificação do odor se inicia no sistema sensor, responsável pela captação ou medição dos estímulos odorantes do ambiente frente à sensibilidade de seus sensores. O sistema sensor pode ser formado por um conjunto de sensores distintos, em que cada elemento mede uma propriedade diferente do composto odorante, ou por um único dispositivo que produz um conjunto de medições e/ou características para cada composto odorante. O sistema sensor pode também ser formado por uma combinação das estruturas descritas anteriormente [Yamazaki, 2001; Keller et al., 1995 e Keller et al., 1994].

Cada composto odorante apresentado ao sistema sensor do nariz artificial produz um estímulo característico, frente aos sensores de detecção. Após a captação destes estímulos ou padrões de odores, e antes que estes sinais sejam apresentados ao sistema de reconhecimento de padrões, faz-se necessária uma etapa de pré-processamento dos dados.

O pré-processamento dos padrões de odores é responsável por eliminar ruídos e normalizar os dados gerados pelo conjunto de sensores, para que sejam tratados de forma mais eficiente na fase de classificação. Pré-processados, os padrões de odores são apresentados ao sistema de reconhecimento de padrões.

O sistema de reconhecimento de padrões é parte de fundamental importância no funcionamento dos narizes artificiais. Ele é responsável por classificar os estímulos odorantes captados pelo sistema sensor [Santos, 2000]. O reconhecimento de padrões em narizes artificiais tem sido proposto a partir de várias abordagens no decorrer da última década [Barbosa et al., 2002; Yamazaki e Ludermir, 2001; Santos, 2000; Di Natale et al., 1995; Gardner et al., 1994; Keller et al., 1994; Sundgren et al., 1991 e Gardner et al., 1990], as técnicas mais relevantes para a classificação de odores serão discutidas nas próximas seções.

A tarefa de reconhecimento de odores não se limita à capacidade do sistema em diferenciar odores. Tais sistemas necessitam, além de detectar a presença do odor no ambiente, associar tais padrões a um odor específico para que obtenham resultados significativos. A Figura 2.1 apresenta um esquema da estrutura básica dos narizes artificiais construídos atualmente [Keller et al., 1995].

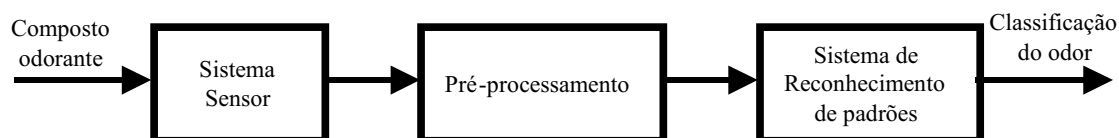


Figura 2.1 – Estrutura básica de um nariz artificial

Nas próximas seções, será apresentada uma breve descrição de cada uma das partes que compõe um nariz artificial.

2.2 Sistema sensor

O sistema sensor dos narizes artificiais é, tal como os neurônios receptores no epitélio olfativo biológico, responsável por converter o estímulo provocado pelos compostos odorantes frente aos seus sensores, em sinais perceptíveis e mensuráveis. O sistema de sensores dos narizes artificiais, em geral, é formado por um conjunto de diversos elementos, com seletividades amplas, possivelmente superpostas e complementares. Existem duas categorias principais de sensores com estas características: *sensores de absorção* e *semicondutores quimicamente sensíveis* [Kress-Rogers, 1997].

2.2.1 Sensores de absorção

O sensor de absorção é formado por uma camada de material absorvente quimicamente seletivo, aplicada em um dispositivo de base acústica ou óptica. Quando em contato com os compostos odorantes, esta camada sofre mudanças em certas propriedades ópticas, como índice de refração, ou mudanças em suas propriedades mecânicas, como

espessura, densidade e elasticidade. Produzindo um sinal relacionado à concentração do odor em contato com a superfície [Santos, 2000].

2.2.2 Semicondutores quimicamente sensíveis

Os sensores quimicamente sensíveis são divididos em duas classes principais: *sensores de semicondutores de óxido metálico* e *sensores de polímeros condutores* [Kress-Rogers, 1997].

2.2.2.1 Semicondutores de óxido metálico

Este sensor é formado por uma película de cerâmica eletricamente aquecida, sobre a qual se deposita um filme fino poroso de óxido metálico, dopado com diversos metais. Os estímulos odorantes interagem com o oxigênio absorvido na superfície, afetando a condutividade do filme de óxido metálico e sua sensibilidade, controlada pela temperatura. Outra forma de modificar as características de resposta deste sensor é através da variação do agente dopante utilizado [Yamazaki, 2001 e Gardner et al., 1994].

2.2.2.2 Polímeros condutores

Os polímeros condutores pertencem a uma família de materiais nos quais, divergindo dos plásticos comuns, podem conduzir corrente elétrica. A condutividade elétrica é possível pela estrutura de ligações químicas duplas e simples, distribuídas alternadamente, nas cadeias moleculares dos polímeros. As propriedades ópticas e elétricas destes plásticos podem ser modificadas pela variação da temperatura, pressão, interação com solventes, ou pela aplicação de um potencial elétrico [Pekker e Janossy, 1986]. Os polímeros condutores formam a base dos sensores do protótipo de nariz artificial [Santos, 2000] utilizado nesta dissertação.

Os sensores construídos com polímeros condutores apresentam diversas vantagens em relação aos demais compostos, entre as quais se destacam [Persaud e Travers, 1997]:

- ❑ a rápida cinética de adsorção em temperatura ambiente,
- ❑ o pequeno consumo de energia (da ordem de microwatts), pois não é necessário um elemento aquecedor,
- ❑ a resistência ao envenenamento por compostos que normalmente tornariam inativos os sensores inorgânicos de semicondutores, como os compostos que contêm enxofre,
- ❑ e a possibilidade de o polímero ser construído com especificidade para determinados compostos químicos.

Exemplos da utilização de sensores de polímeros condutores em narizes artificiais podem ser encontrados em [Santos, 2000 e Gardner et al., 1994].

2.3 Pré-processamento

A fase de pré-processamento dos dados tem como finalidade a análise dos dados, extração dos descritores, normalização e remoção ou diminuição dos erros na precisão das análises posteriores a que serão submetidos os sinais de odores. Esta fase consiste da realização de operações sobre os sinais captados pelo sistema sensor, a fim de corrigir distorções existentes nos dados ou melhorar o desempenho de algoritmos de classificação que serão utilizados no reconhecimento dos odores.

Existem muitos métodos de pré-processamento de dados, eles normalmente se caracterizam pelo resultado obtido em sua utilização. Na escolha do método de pré-processamento para o tratamento dos sinais de odores, algumas características inerentes ao processo de aquisição dos dados devem ser consideradas: (1) os sensores utilizados para captar os sinais de odores podem apresentar distúrbios ou ruídos; (2) o processo de aquisição dos dados tende a ser instável, ou seja, os sensores apresentam variações durante o processo de aquisição dos dados, as quais podem se transformar em *outliers*⁶ na base de dados; e (3) a propagação do sinal através do canal de comunicação entre os

⁶ *Outliers* são observações que não seguem a distribuição da maioria dos dados.

sistemas sensor e de reconhecimento de padrões pode ser contaminada por sinais de interferência originados do ambiente.

Outra questão a considerar no pré-processamento é quanto a normalização dos padrões, o ideal é que os dados estejam compreendidos entre intervalos bem definidos. A normalização se faz necessária para que valores esparsos não interfiram no processo de aprendizado e convergência do algoritmo utilizado no reconhecimento dos sinais de odores. As características do classificador utilizado também podem influenciar na escolha do método de pré-processamento.

Para o tratamento de sinais de odores em narizes artificiais, diversos modelos de normalização já foram propostos, dentre os quais podem ser destacados [Gardner e Bartlett, 1994]:

□ Modelo da Diferença:

$$r_{ij} = x_{ij}^{\max} - x_{ij}^{\min}, \quad (2.1)$$

□ Modelo Relativo:

$$r_{ij} = \frac{x_{ij}^{\max}}{x_{ij}^{\min}}, \quad (2.2)$$

□ Modelo de Diferença Fracional:

$$r_{ij} = \frac{x_{ij}^{\max} - x_{ij}^{\min}}{x_{ij}^{\min}}, \quad (2.3)$$

□ e Modelo Logarítmico:

$$r_{ij} = \log(x_{ij}^{\max} - x_{ij}^{\min}), \quad (2.4)$$

onde $x_{ij}^{\max}$ e $x_{ij}^{\min}$ são, respectivamente, os valores de máximo e mínimo dos sinais captados pelos sensores do nariz artificial.

Sendo assim, a resposta dos sensores a um conjunto de q odores pode ser representada por uma matriz R , em que cada coluna j é a resposta do conjunto de sensores ao odor i :

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1q} \\ r_{21} & r_{22} & \cdots & r_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pq} \end{bmatrix} \quad (2.5)$$

A complexidade do problema de análise de padrões está relacionada com o número de termos não-nulos da matriz R [Gardner e Bartlett, 1994]. Por exemplo, se os p sensores forem independentes, ou seja, se cada sensor responder apenas a um entre os q odores, então cada linha só possuirá um elemento não-nulo. Neste caso, a matriz de resposta é uma maneira conveniente de manipular os dados, não sendo necessário utilizar métodos mais sofisticados de análise de padrões. Em geral, o conjunto de sensores não é independente, de modo que cada sensor responde a uma variedade de odores com sensibilidade diferente. Dessa forma, uma técnica de reconhecimento de padrões mais elaborada é necessária para processar os sinais originados pelo conjunto de sensores.

Assim, ruídos, distorções, compressão de sinal e normalização dos dados devem ser considerados pelos mecanismos de pré-processamento utilizados em narizes artificiais, de forma a facilitar o processo de reconhecimento dos estímulos odorantes.

2.4 Sistemas de reconhecimento de padrões para narizes artificiais

Em geral, os sistemas de reconhecimento de padrões são divididos em paramétricos e não-paramétricos [Duda et al., 2001], a escolha de um determinado sistema de reconhecimento, classificador ou regra de decisão depende de fatores como, por exemplo, a qualidade dos dados disponíveis e o resultado desejado na classificação.

Sistemas de reconhecimento de padrões paramétricos partem do pressuposto que os dados observados (para cada classe), podem ser descritos por uma função densidade de probabilidade. Normalmente, admite-se que os dados possuam uma distribuição normal, com média e variância conhecidas. Este tipo de técnica exige a construção de

uma base de conhecimento, contendo as funções densidade de probabilidade apropriadas, para que sejam construídas as regiões de decisão. Enquanto que métodos não-paramétricos não fazem tais suposições, eles utilizam funções discriminantes para determinarem as regiões de decisão.

As principais técnicas de reconhecimento de padrões utilizadas em narizes artificiais serão brevemente comentadas a seguir.

2.4.1 Análise de Função Discriminante

A Análise de Função Discriminante [Tabachnick e Fidell, 1996] é um método paramétrico, cujo objetivo consiste em identificar as variáveis que melhor discriminam agrupamentos de dados previamente fixados. Os grupos de dados bem definidos são fixados como variáveis dependentes (ou explicadas), enquanto que as variáveis do problema são identificadas como variáveis independentes (ou explicativas). A Análise de Função Discriminante assume que os dados possuem distribuição multi-normal e determina funções discriminantes relacionadas linearmente com os vetores dos dados a serem analisados.

Exemplos de utilização da Análise de Função Discriminante em narizes artificiais podem ser encontrados em [Gardner e Bartlett, 1992 e Gardner et al., 1992].

2.4.2 Análise de Agrupamentos

A Análise de Agrupamentos [Tabachnick e Fidell, 1996], é um método não-supervisionado, que objetiva agrupar os dados, de forma a permitir a identificação de semelhanças entre os padrões. Existem diversas abordagens para Análise de Agrupamentos, na resolução de problemas de classificação se busca uma forma geral de maximizar a semelhança inter grupo e a dessemelhança entre grupos.

Alguns exemplos da utilização da análise de agrupamentos na classificação de padrões de odores em narizes artificiais podem ser vistos em [Gardner e Bartlett, 1992 e Gardner, 1991].

2.4.3 Análise de Componentes Principais

A Análise de Componentes Principais [Duda et al., 2001], também conhecida como *Transformação de Karhunen-Loeve*, é um método estatístico multivariado simples, não-paramétrico, não-supervisionado e linear. A Análise de Componentes Principais pode ser utilizada para compressão de dados, redução de dimensionalidade, extração de características e projeção de dados em conjuntos multivariados. Maiores informações sobre esta técnica podem ser encontradas no Capítulo 3 e no Apêndice A desta dissertação.

Aplicações da Análise de Componentes Principais em narizes artificiais podem ser encontrados em [Di Natale et al., 1995; Gardner e Bartlett, 1992 e Gardner, 1991].

2.4.4 Redes Neurais Artificiais

As redes neurais artificiais são modelos não-paramétricos e, em geral, não-lineares, compostos por unidades de processamento paralelamente interconectadas, normalmente adaptativas, cuja organização é baseada em modelos físicos de sistemas biológicos [Braga et al., 2000 e Hayken, 1994]. As redes neurais artificiais constituem uma arquitetura funcional que estabiliza os estímulos sensoriais percebidos em suas entradas, fixam o aprendizado através da construção de uma topologia plástica e mutante, desenhada pelas conexões entre as unidades de processamento criadas no processo de aprendizagem. Desta forma, as redes neurais artificiais possuem a capacidade de adquirir sua base de conhecimento (aprender) através de exemplos e de fazer interpolações e extrapolações desta base de conhecimento.

A utilização de redes neurais artificiais na análise de padrões de odores tem apresentado resultados promissores. Em Santos (2000), são apresentados três protótipos de narizes artificiais, todos com sensores a base de polímeros condutores. O primeiro protótipo é composto por três sensores distintos, tendo a finalidade de classificar substâncias simples (Metanol, Tetracloreto de Carbono e Etanol), através da aplicação de uma rede neural Perceptron Multi-Camadas, com uma camada escondida, treinada com algoritmo Backpropagation [Rumelhart et al., 1986]. O segundo protótipo

apresenta melhorias na construção dos sensores e na aquisição dos dados, sendo testado para reconhecer odores de substâncias mais complexas, como diferentes safras de vinho. Foram aplicadas redes Perceptron Multi-Camadas, Função de Base Radial [Bishop, 1986] e Redes de Elman [Braga et al., 2000], permitindo uma comparação do desempenho das diferentes abordagens conexionistas. Os resultados obtidos na classificação dos odores não apresentaram diferenças significativas entre as três abordagens utilizadas. O terceiro protótipo apresenta mudanças na tecnologia utilizada para produzir os sensores. O modelo foi testado com substâncias puras (Etanol, Benzeno, Metanol e Tetracloro de Carbono) e com algumas misturas destas substâncias (Etanol com Tetracloro de Carbono e Benzeno com Metanol). Também foram realizados testes com dados, provenientes de três safras distintas de um mesmo vinho comercial e experimentos para distinguir vinho de whisky.

Outro trabalho, Barbosa et al. (2002) apresenta os resultados da implementação e testes de uma rede neural Função de Base de Radial, como sistema de reconhecimento de padrões de odores. Esta rede foi utilizada como classificador em um nariz artificial, formado pelo arranjo de oito sensores de polímeros condutores [Santos, 2000]. Este protótipo foi utilizado na classificação de uma base de dados composta por gases derivados da indústria petrolífera, os testes de classificação foram realizados com cinco gases distintos (Etano, Metano, Butano, Propano e Monóxido de Carbono). No mesmo trabalho, são realizados testes com diversas abordagens de pré-processamento de odores, resultando em significativa melhora na classificação dos estímulos odorantes.

Em Yamazaki (2001) é apresentada uma abordagem que considera as características temporais dos sinais gerados pelos sensores do nariz artificial, na classificação dos padrões de odores. Yamazaki (2001) demonstra a partir de experimentos comparativos, que a abordagem temporal mostra-se mais adequada para a classificação de padrões de odores, quando comparada a redes do tipo Perceptron Multi-Camadas. Neste trabalho, o autor utilizou uma Rede Neural com Atraso no Tempo para o reconhecimento dos padrões de odores. Os testes comparativos foram realizados com uma base de dados proveniente de três safras distintas de um mesmo vinho tinto

comercial. A base de dados foi obtida a partir do protótipo de um nariz artificial [Santos, 2000], formado por um conjunto de seis sensores de polímeros condutores.

2.5 Sumário do capítulo

Este capítulo apresentou uma descrição teórica sobre os narizes artificiais. Foram discutidas as principais etapas envolvidas no funcionamento de um nariz artificial: o sistema de sensor, a fase de pré-processamento dos dados e o sistema de reconhecimento de padrões. No decorrer do capítulo foram descritas cada uma das etapas do reconhecimento de odores em um nariz artificial, as principais partes que as constituem e relacionadas as principais técnicas utilizadas na literatura para o reconhecimento de estímulos odorantes.

No próximo capítulo será abordada a base de padrões de odores utilizada nesta dissertação. Uma investigação da base de dados será realizada, a partir de uma técnica estatística multivariada, a fim de verificar algumas características do problema abordado.

3

Descrição do Problema e Análise dos Dados

Neste capítulo é apresentada uma descrição da base de padrões de odores utilizada nesta dissertação. Os padrões de odores são descritos e posteriormente analisados com uma técnica de análise multivariada. A seção 3.1 trata sobre a aquisição dos padrões de odores pelo nariz artificial. Na seção 3.2 é apresentada a técnica de Análise de Componentes Principais, utilizada como método de visualização e interpretação dos sinais de odores. A seção 3.3 apresenta um sumário do capítulo.

3.1 Aquisição dos dados

O processo de aquisição dos estímulos odorantes pelo nariz artificial é realizado no sistema sensor, o qual é composto por uma matriz de sensores que captam informações do ambiente e as convertem em medidas quantitativas de interesse.

Vários tipos de sensores e dopantes característicos podem ser utilizados no sistema sensor do nariz artificial. Cada sensor do nariz artificial é responsável por uma resposta específica a cada odor analisado. Ou seja, todo composto odorante apresentado ao sistema sensor produz nos sensores alterações físicas ou químicas que o caracterizam.

A base de dados de odores, utilizada nesta dissertação, foi obtida a partir da exposição do protótipo de um nariz artificial [Santos, 2000], frente a compostos fornecidos pela companhia petrolífera Petrobras⁷. A partir deste protótipo foram realizados testes de sensibilidade no arranjo de oito sensores baseados em *filmes de polipirrol*⁸. Os sensores foram preparados com diferentes dopantes, por polimerização *in situ*⁹, em *substratos de ITO* (substrato vítreo condutor). Os testes foram realizados frente aos gases derivados do petróleo: Metano, Etano, Propano e Butano. Também foi verificada a sensibilidade frente ao gás Monóxido de Carbono.

A aquisição dos padrões de odores no nariz artificial envolveu três etapas distintas:

- a) Coleta da linha de base (com duração de 10 minutos), onde os sensores são expostos a um gás inerte (gás Nitrogênio);
- b) Coleta dos valores de resistência do arranjo de sensores, frente ao gás analisado (com duração de 40 minutos); e
- c) Purga ou limpeza dos sensores com gás Nitrogênio (com duração média de 5 minutos).

Os sensores utilizados na aquisição dos dados são denominados pelo nome do dopante utilizado em sua constituição. Nesta análise, o arranjo de sensores foi formado pelos sistemas apresentados na Tabela 3.1.

Tabela 3.1 - Descrição dos sensores utilizados no nariz artificial

Sensor	Dopante	Nome do dopante
01	DBS	Dodecilbenzoato de Sódio
02	OSA	Octil-sulfonato de Sódio
03	NBS	Ácido Nitrobenzeno Sulfônico
04	ASA	Antraquinona Sulfonato de Sódio
05	PTSA	Ácido P-toluenosulfônico
06	CAS	Ácido Canforsulfônico
07	PER	Perclorato de Lítio
08	OSA2	Octil-sulfonato de Sódio

⁷ Petróleo Brasileiro S.A., Brasil.

⁸ Polímero Condutor, material orgânico, do tipo "plástico", geralmente derivado do petróleo, que conduz eletricidade.

⁹ No local de origem; de forma natural.

A partir da configuração de sensores descrita, os dados foram obtidos com nove repetições de testes, para os cinco gases analisados. Em todas as repetições, para os cinco gases, o valor de resistência de cada sensor foi registrado a cada vinte (20) segundos. Como o processo de aquisição dos padrões de cada arranjo estendeu-se por quarenta (40) minutos, cada sensor obteve o registro de 120 valores, para cada um dos cinco gases apresentados ao nariz artificial.

O conjunto formado pelos oito valores, de cada um dos sensores no mesmo instante de tempo, foi considerado uma instância da base de dados. Desta forma, cada repetição contém 600 padrões, formados por 960 valores de cada gás. A partir das nove repetições, obteve-se um total de 5.400 padrões que formam a base de dados.

A Tabela 3.2 apresenta um exemplo de uma das repetições de aquisição dos dados, para o gás Butano, permitindo uma melhor visualização dos padrões que estão sendo tratados.

Tabela 3.2 - Exemplo de aquisição dos dados para o gás Butano

Tempo (s)	Sensores – valores de resistência							
	DBS	OSA	NBS	ASA	PTSA	CAS	PER	OSA2
0,20	68400	360	614	141,6	269,2	32220	32380	498
0,40	68600	360	614	141,6	269,4	32280	32360	498
1,00	68600	360	614	141,6	269,4	32220	32340	498
1,20	68800	360	614	141,6	269,4	32280	32360	500
1,40	68800	360	614	141,6	269,4	32300	32320	500
...	...	...	...	...	...	...	...	...
40,00	69200	360	618	141,8	270,8	32500	31820	500

Como o processo de aquisição dos dados é realizado de forma contínua no tempo, cada sensor gera uma curva temporal, contendo os valores de resistência dos gases analisados. A Figura 3.1 apresenta um exemplo da evolução temporal dos valores de resistência, para o Sensor 1 (*Dodecilbenzoato de sódio*), durante a fase de aquisição dos estímulos do gás Butano. Pode-se perceber uma grande oscilação nos valores mensurados em um pequeno espaço de tempo. Porém, maiores análises seriam necessárias para verificar as características temporais desta série, as quais não estão entre os objetivos desta dissertação.

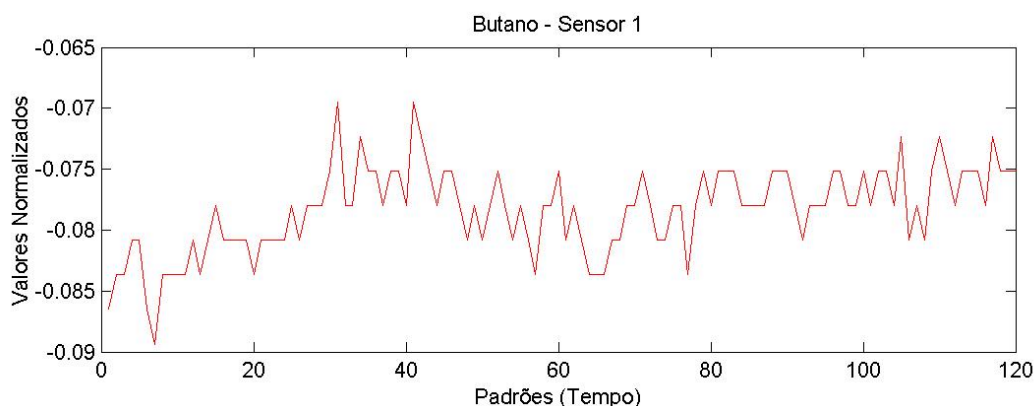


Figura 3.1- Curva temporal formada pelos valores de resistência do gás Butano

3.2 Análise dos dados

Um sistema de multi-sensores, como o sistema sensor do nariz artificial, produz dados com grande dimensionalidade. Ou seja, um grande número de variáveis que caracterizam ou descrevem cada observação. A visualização simultânea de dados em mais de três dimensões é muito difícil de ser analisada. Assim, métodos para reduzir a dimensionalidade de séries de dados multivariados são necessários, pois possibilitam a análise de bases de dados de grande magnitude e complexidade.

A Análise Multivariada de dados corresponde a um conjunto de métodos estatísticos em que os objetos em estudo são caracterizados por uma multiplicidade de variáveis. Estes métodos permitem a análise simultânea de alterações em várias propriedades que caracterizam os dados investigados. Existem várias técnicas clássicas para análise de dados multivariados, sua utilização depende do tipo de dado que se deseja analisar e dos objetivos do estudo.

Dentre as técnicas empregadas na análise multivariada de dados, neste estudo, foi utilizada a *Análise de Componentes Principais* [Jackson, 1991]. Esta técnica é um método estatístico multivariado simples, que pode ser utilizado para compressão, redução da dimensionalidade, extração de características e projeção de dados multivariados. Maiores detalhes sobre a técnica de Análise de Componentes Principais podem ser encontrados no Apêndice A do presente trabalho.

Nos próximos parágrafos são apresentados os resultados da aplicação da Análise de Componentes Principais na base de dados de padrões de odores.

Um procedimento básico para melhorar a análise dos dados, de modo a reduzir a dimensão das amostras estudadas, é a escolha do número de componentes principais (CPs) utilizados na descrição do sistema. Esta escolha é realizada considerando a porcentagem de variância descrita pelas CPs e a variância residual dos dados. Na Tabela 3.3 são apresentadas as componentes principais que descrevem os dados, onde *var* corresponde a variância descrita, *var (%)* corresponde a percentagem de variância explicada e *ac* diz respeito aos seus respectivos valores acumulados.

Tabela 3.3 – Variância descrita pelas Componentes Principais

<i>Componentes Principais</i>	<i>var</i>	<i>var (%)</i>	<i>av(ac)</i>	<i>var(% ac)</i>
CP1	7.0805	88.5062	7.0805	88.5062
CP2	0.7826	9.7826	7.8631	98.2888
CP3	0.0988	1.2349	7.9619	99.5237
CP4	0.0206	0.2569	7.9825	99.7806
CP5	0.0086	0.1074	7.9911	99.8880
CP6	0.0065	0.0808	7.9976	99.9688
CP7	0.0015	0.0187	7.9991	99.9875
CP8	0.0010	0.0125	8.0001	100.0000

A partir da Tabela 3.3 se pode verificar que a CP1 descreve somente 88.5% da variância total dos dados, o que é insuficiente para descrever o conjunto total dos dados. Isto se deve pelo fato da variância residual ainda ser muito alta; desta forma alguma informação poderá estar sendo perdida. Por exemplo, a contribuição de variáveis com alto *peso* na CP2 poderia estar sendo desprezada, o que prejudicaria a descrição dos dados. Faz-se necessário utilizar mais uma componente na descrição dos dados. A CP2 descreve 9.78% da variância total dos dados e somada à CP1, descrevem 98.28% da variância total descrita.

Apesar da quantidade de variância descrita pela CP3 (apenas 1.32%) ser pequena, sua inclusão entre as componentes analisadas pode ser interessante, somadas, as três CPs correspondem a mais de 99.5% da variância total dos dados analisados. A relação entre a variância explicada em cada uma das componentes principais pode ser

facilmente visualizada na Figura 3.2. Desta forma, nesta análise, iremos considerar as três primeiras componentes principais dos dados.

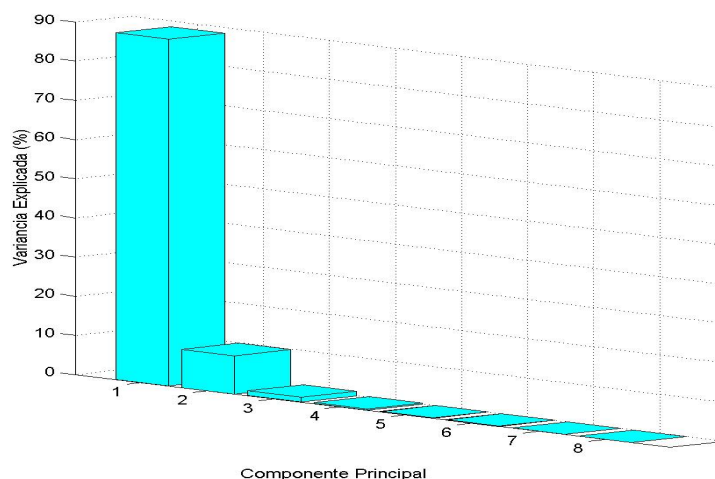


Figura 3.2 – Variância explicada em cada uma das componentes principais

A projeção de cada padrão de odor i ($i = 1, 2, \dots, n$) neste novo sistema de eixos, fornece os *escores*¹⁰; cada amostra possui um valor de *escore* para cada um dos novos eixos (as CPs). A projeção destes valores, nas três primeiras componentes principais, pode ser vista na Figura 3.3; a partir deste gráfico pode-se observar que a distinção entre a maioria dos gases é bem definida, apenas os gases Metano e Etano mantêm-se agrupados em um plano de difícil separação.

Na Figura 3.3 se pode observar também que os agrupamentos dos gases Metano, Etano e Monóxido de Carbono apresentam amostras que aparecem separadas de seus respectivos agrupamentos. Esta separação pode caracterizar a presença de *outliers* ou ruídos nos dados, os quais podem dificultar a fase de classificação dos odores, dependendo do tipo de classificador utilizado. Estas amostras, que apresentaram algumas diferenças em relação aos padrões de seus agrupamentos, serão mantidas na base de dados para que sirvam de parâmetro no teste dos classificadores que serão utilizados nesta dissertação.

¹⁰ O termo em inglês *scores* também é muito utilizado na literatura sobre o tema.

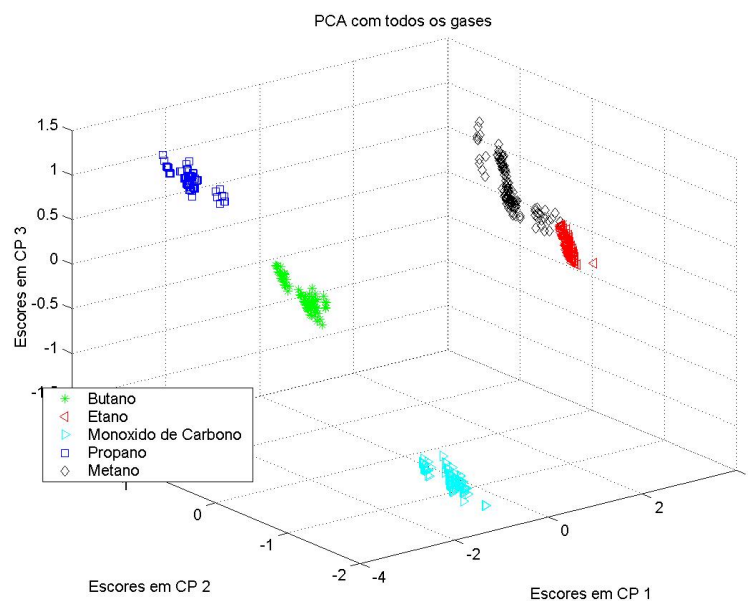


Figura 3.3 - Projeção dos dados em três componentes principais

A Análise de Componentes Principais também pode ser utilizada para verificar a importância de cada uma das variáveis do conjunto original de dados. Ou seja, as variáveis originais ou os sensores em nosso caso, com maior *peso* na combinação linear dos primeiros componentes principais são os mais importantes do ponto de vista estatístico.

O resultado da análise dos *pesos* de cada uma das variáveis é apresentado na Tabela 3.4. A partir desta tabela e da Figura 3.4, pode-se verificar quais sensores mais contribuem para a separação dos gases em cada uma das componentes principais. A análise pode ser feita primeiramente utilizando a Figura 3.4, de forma a identificar quais das variáveis possuem menor correlação, a partir destas variáveis pode-se identificar na Tabela 3.4 os sensores com maior relevância em cada uma das componentes principais. Desta forma, podemos verificar que quatro sensores (DBS, ASA, PER e OSA2) possuem maior relevância para a separação dos gases nas componentes principais analisadas.

Tabela 3.4 – Peso de cada variável nas Componentes Principais

<i>Sensor</i>	<i>Componente Principal</i>							
	<i>CP1</i>	<i>CP2</i>	<i>CP3</i>	<i>CP4</i>	<i>CP5</i>	<i>CP6</i>	<i>CP7</i>	<i>CP8</i>
DBS	-0.2780	-0.7363	0.5234	0.2502	0.1879	-0.0485	0.0307	-0.0743
OSA	-0.3388	0.4717	0.1255	0.7142	0.1413	0.3407	-0.0267	0.0064
NBS	-0.3661	0.2441	0.1493	-0.0987	-0.1859	-0.4123	0.7174	-0.2347
ASA	0.3535	0.2597	0.7800	-0.2274	-0.2553	0.2869	-0.0037	-0.0025
PTSA	-0.3734	0.0944	0.1438	0.0190	-0.4884	-0.4389	-0.5040	0.3813
CAS	-0.3670	0.2123	0.1579	-0.4906	0.6562	0.0280	-0.0606	0.3458
PER	0.3665	0.2283	0.1819	0.2038	0.4116	-0.6310	-0.2614	-0.3210
OSA2	-0.3749	0.0475	-0.0318	-0.2885	-0.0797	0.1940	-0.3970	-0.7559

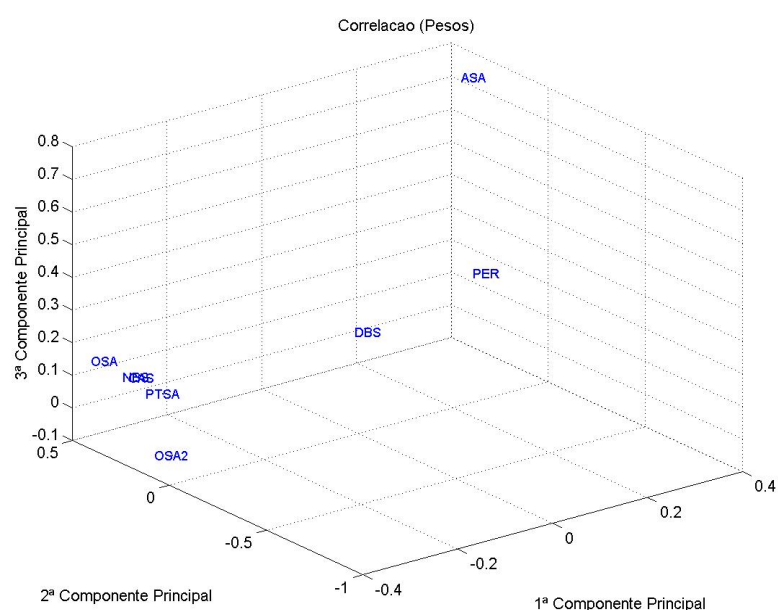


Figura 3.4 - Correlação entre os sensores

3.3 Sumário do Capítulo

Este capítulo teve como objetivo apresentar uma descrição da base de dados utilizada nesta dissertação e sua forma de aquisição frente aos sensores do nariz artificial. Foi apresentada também uma análise dos padrões de odores a partir da técnica de Análise de Componentes Principais.

O interesse na utilização da Análise de Componentes Principais reside no fato desta técnica permitir estabelecer associações entre os gases e os sensores, de acordo

com suas propriedades de distribuição estatística, evidenciadas por este método de manipulação estatística. A utilização da Análise de Componentes Principais também é importante por ser um método que proporciona a visualização da distribuição espacial dos dados, mesmo que de forma simplificada. Isto porque os componentes principais são aproximações com menores dimensões, das variáveis originais.

A partir dos resultados da Análise de Componentes Principais e da visualização da distribuição dos padrões de odores nas três primeiras componentes principais, pode-se verificar que a maioria dos gases apresenta grande separabilidade espacial. Apenas os gases Metano e Etano apresentam características estruturais muito próximas, as quais podem dificultar sua diferenciação durante a fase de classificação. Vale ressaltar que na transposição dos dados para o plano das componentes principais houve uma simplificação da distribuição dos mesmos, fato pelo qual aparecem tão espacialmente disjuntos.

Outro ponto verificado na análise é que apenas alguns sensores apresentam maior relevância na separação dos gases. Quatro sensores demonstraram possuir maior relevância na classificação dos odores, são eles: Dodecilbenzoato de Sódio, Antraquinona Sulfonato de Sódio, Perclorato de Lítio e Octil-sulfonato de Sódio.

Na sequência serão apresentadas as técnicas conexionistas e híbridas utilizadas como sistemas de reconhecimento de padrões nesta dissertação.

4

Sistemas de Reconhecimento de Padrões para Narizes Artificiais

Neste capítulo são apresentadas as técnicas utilizadas, nesta dissertação, como sistemas de reconhecimento de sinais de odores. Na seção 4.1 são apresentadas às abordagens conexionistas de reconhecimento de padrões utilizadas, as técnicas descritas são: Redes Neurais Perceptron Multi-Camadas e Redes Neurais com Atrasos no Tempo. A seção 4.2 descreve as abordagens neurais híbridas propostas para a classificação de padrões de odores, são elas: as Redes Feature-weighted Detector e as Redes Evolving Fuzzy Neural Networks. Na seção 4.3 é apresentado um sumário do capítulo.

4.1 Abordagens conexionistas para reconhecimento de padrões

Conexionismo denota uma forma particular de processamento de informação, sendo já há algum tempo, uma área de pesquisa que envolve indistintamente várias disciplinas do conhecimento. Os sistemas conexionistas são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano. Caracterizam-se pela reunião de uma grande quantidade de células (nodos ou unidades) de processamento, interligadas por um grande número de conexões, que processam a informação de forma paralela.

Os sistemas conexionistas ou redes neurais artificiais têm sido bastante atrativos, pelas funções que desempenham no reconhecimento de padrões e em outras áreas da Inteligência Computacional. As redes neurais artificiais possuem a capacidade de armazenar informações e torná-las disponíveis para uso posterior. Assim como no cérebro humano, o conhecimento é adquirido do meio externo e consolidado por um processo de aprendizado ou treinamento.

No treinamento das redes neurais artificiais, as informações acerca do ambiente ou processo em que se deseja tratar são representadas através de vetores de informações. Estes vetores representam as amostras de dados disponíveis e são as entradas das redes neurais artificiais, podendo ser rotuladas ou não, caso a classificação das amostras seja conhecida ou não, respectivamente.

Nas redes neurais artificiais, o conhecimento é armazenado nas conexões entre as unidades de processamento, ou pesos, que em seu correspondente biológico são conhecidas como sinapses. O algoritmo de aprendizado modifica a intensidade das sinapses, de forma a alcançar o objetivo desejado. Em alguns casos, o aprendizado pode modificar a topologia da rede [Braga et al., 2000].

Como exposto na seção 2.4.4 desta dissertação, as redes neurais artificiais contribuíram amplamente para a evolução dos sistemas de reconhecimento de padrões dos narizes artificiais. Nas seções seguintes são descritas duas das principais abordagens conexionistas para o reconhecimento de sinais de odores.

4.1.1 Redes Neurais Perceptron Multi-Camadas

Dentre os vários modelos de redes neurais artificiais existentes, a rede *Perceptron Multi-Camadas* [Rumelhart et al., 1986] é a rede mais difundida, parte por sua facilidade de implementação, parte por sua simplicidade. Características estas, que a tornaram uma das redes neurais artificiais mais utilizadas na classificação de padrões de odores [Yamazaki, 2001; Santos, 2000; Di Natale et al., 1995; Gardner et al., 1994; Keller et al., 1994; Sundgren et al., 1991 e Gardner et al., 1990].

As redes Perceptron Multi-Camadas utilizam unidades de processamento do tipo *McCulloch e Pitts* (1943), com funções de ativação normalmente não-lineares, organizadas em várias camadas. A Figura 4.1 apresenta uma rede Perceptron Multi-Camadas com 3 níveis, sem laços de realimentação (conhecida como arquitetura de propagação ou *feedforward*).

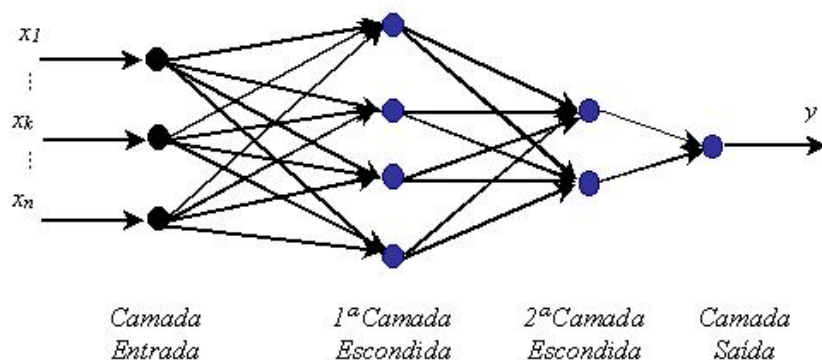


Figura 4.1 - Exemplo de uma rede Perceptron Multi-Camadas

As principais características de uma rede Perceptron Multi-Camadas são:

- ❑ Contém uma ou mais camadas escondidas.
- ❑ O modelo de cada unidade de processamento inclui uma função de ativação não-linear, geralmente é adotada a função logística (sigmóide) ou a função tangente hiperbólica.
- ❑ Apresenta um alto grau de conectividade entre as camadas.
- ❑ Utiliza um algoritmo com retropropagação no ajuste das conexões entre as unidades de processamento.

Existem vários algoritmos para treinar redes Perceptron Multi-Camadas. Nos experimentos realizados nesta dissertação, foi utilizado o algoritmo de treinamento *Levenberg-Marquardt* [Lang e Hilton, 1988]. Este algoritmo é uma variação do algoritmo *Backpropagation* [Rumelhart et al., 1986], normalmente utilizado para acelerar o tempo de treinamento da rede neural. Sua principal diferença em relação ao algoritmo *Backpropagation*, é que o *Backpropagation* utiliza gradiente descendente como método de aproximação do mínimo da função de erro, enquanto que o algoritmo

de *Levenberg-Marquardt* utiliza uma aproximação pelo método de Newton [Hagan e Menhaj, 1994].

4.1.2 Redes Neurais com Atrasos no Tempo

A solução para muitos problemas na área de Inteligência Computacional exige arquiteturas de redes neurais artificiais que sejam capazes de realizar mapeamentos dinâmicos sobre os dados. Para que uma rede neural artificial seja considerada dinâmica, esta deve possuir memória [Elman, 1990]. A inclusão de memória nas redes neurais artificiais pode ser feita de duas formas: (1) se a rede considerar entradas atrasadas no tempo; ou (2) se a rede possuir laços de realimentação.

Este trabalho tem particular interesse em *Redes Neurais com Atraso no Tempo*, a qual, é a arquitetura conexionista que tem apresentado os resultados mais promissores na classificação de padrões de odores [Zanchettin e Ludermir, 2003b; Zanchettin et al., 2003; Yamazaki, 2001; Yamazaki e Ludermir, 2001 e Yamazaki et al., 2001].

A Rede Neural com Atraso no Tempo (*Time Delay*), foi proposta por Waibel et al. (1989), com a finalidade de resolver o problema de reconhecimento de fonemas isolados utilizando um espectrograma. Em essência, uma Rede Neural com Atrasos no Tempo é uma rede neural artificial com arquitetura *feedforward*, onde a resposta da rede no tempo t é baseada nas entradas no tempo $(t-1)$, $(t-2)$, ..., $(t-n)$. Com isso, considera-se um histórico da seqüência temporal dos dados.

A abordagem, onde os atrasos são fornecidos somente à camada de entrada é muito utilizada, denominada “*janelamento*”, introduz memória somente nas unidades de processamento da primeira camada escondida. A Figura 4.2, ilustra uma rede com duas camadas escondidas utilizando o método de *janelamento*.

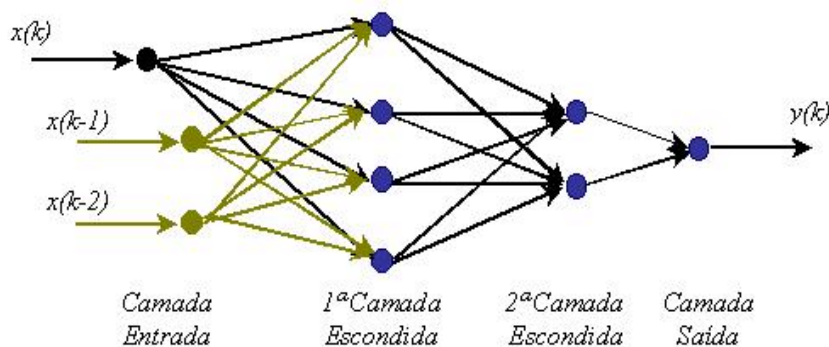


Figura 4.2 - Método de *janelamento* para processamento temporal

As unidades de processamento da primeira camada escondida, além de receberem a entrada atual $x(k)$, recebem também, neste caso, as duas entradas anteriores: $x(k-1)$ e $x(k-2)$; criam-se, porém, oito novas conexões entre as unidades de processamento da camada de entrada e camada escondida. Ou seja, o método de *janelamento* proporcionou à rede uma *memória de ordem 2*, só na primeira camada escondida. Apesar de introduzir memória ao modelo, esta arquitetura pode ser considerada um caso particular de uma rede estática padrão, totalmente interconectada. Pois, as novas unidades de processamento são consideradas totalmente independentes das entradas originais.

Existem ainda generalizações do método para considerar atrasos também nas camadas escondidas e na saída da rede. A Figura 4.3 ilustra um exemplo de outro modelo de Rede Neural com Atraso no Tempo, denominada rede *Time Delay Neural Network* [Lang e Hilton, 1988]. Esta rede proporciona memória a todos os neurônios da(s) camada(s) escondida(s) e da camada de saída.

A rede apresentada na Figura 4.3 contém somente uma camada escondida (podendo apresentar mais de uma) com *memória de ordem 1*, e a camada de saída possui *memória de ordem 2*. A primeira camada escondida além da entrada $x(k)$ recebe também uma entrada temporalmente anterior $x(k-1)$, já a camada de saída recebe duas entradas temporalmente anteriores para cada uma das unidades escondidas, $a_1(k-1)$, $a_2(k-1)$, $a_3(k-1)$, $a_1(k-2)$, $a_2(k-2)$ e $a_3(k-2)$. De maneira similar ao método de *janelamento*, existem novas conexões entre as unidades de processamento, mas a rede

Time Delay Neural Network não conduz à forma de uma rede estática padrão totalmente interconectada. A diferença entre as duas abordagens se deve ao fato que as novas unidades de processamento da camada escondida da rede *Time Delay Neural Network* não são conectadas aos elementos de entrada.

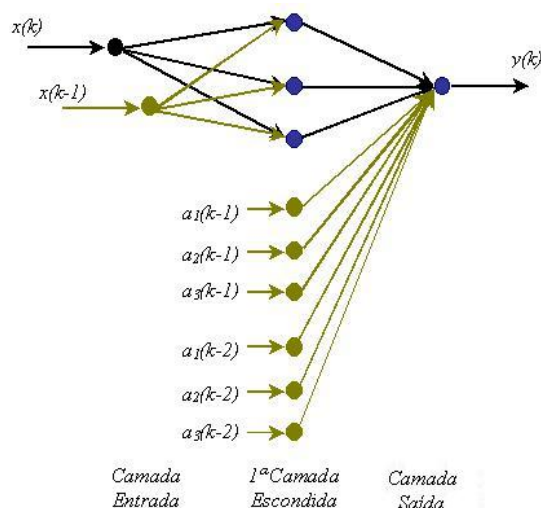


Figura 4.3 - Rede Time Delay Neural Network para processamento temporal

De maneira geral, as Redes Neurais com Atrasos no Tempo apresentam as seguintes propriedades:

- ❑ Múltiplas camadas com interconexão suficiente entre as unidades de processamento e entre cada uma destas camadas.
- ❑ Capacidade de representar relacionamentos entre os eventos temporais.
- ❑ As características reais ou abstrações da aprendizagem da rede não variam no tempo.
- ❑ O processo de aprendizagem não requer um alinhamento temporal preciso dos rótulos que serão aprendidos.

4.2 Abordagens neurais híbridas para o reconhecimento de padrões

O termo *Sistema Híbrido* é utilizado quando se faz referência a uma combinação de diferentes técnicas e modelos de implementação. Dentro da Inteligência Computacional, um Sistema Híbrido Inteligente é aquele que utiliza um paradigma inteligente em combinação com alguma outra técnica computacional, seja ela inteligente ou não. A combinação de um sistema especialista com alguma técnica de otimização, por exemplo, também pode ser enquadrada nessa definição.

A integração de duas ou mais técnicas computacionais, procura combinar vantagens, com o intuito de superar as desvantagens que cada técnica apresenta individualmente, na resolução de problemas complexos e computacionalmente exigentes.

Abordagens híbridas também possibilitam a criação de sistemas com múltiplas capacidades de processamento, agregados em uma única arquitetura. Se um problema pode ser dividido em módulos de processamento independentes, então vários módulos inteligentes podem ser utilizados para gerar soluções parciais para este problema. Sendo que estes módulos podem corresponder a abordagens computacionais diferentes.

É possível, por exemplo, usar uma rede neural artificial, um sistema especialista e um algoritmo genético de maneira competitiva ou cooperativa, para a solução de um problema que envolva reconhecimento de padrões, tomada de decisão e otimização.

Os sistemas *Neuro-Difusos* são a abordagem híbrida investigada nesta dissertação, são constituídos a partir da integração de técnicas conexionistas e conjuntos difusos [Abraham e Nath, 2000]. As redes neurais artificiais, como previamente enfatizado, são métodos computacionais que foram desenvolvidos tendo como inspiração o desempenho do cérebro humano. A teoria dos conjuntos difusos é, por sua vez, capaz de modelar a incerteza e a ambigüidade do mundo real. Os conjuntos difusos são especialmente adequados na descrição de sistemas de processamento de informação complexos, não-lineares ou não claramente definidos.

Uma integração criteriosa das redes neurais artificiais e dos conjuntos difusos pode então resultar em sistemas mais inteligentes (em termos de paralelismo, não-linearidade, tolerância a faltas, adaptabilidade, gerenciamento da incerteza, modelagem e controle de processos dinâmicos), que se aproximam mais da forma como os humanos trabalham [Abraham e Nath, 2000].

Várias são as formas de integração das redes neurais artificiais e dos sistemas difusos. Por exemplo, a nebulosidade do sistema difuso pode ser implementada, a partir das redes neurais artificiais, na entrada do sistema, na função de decisão de saída (função de pertinência) ou na estratégia de aprendizagem [Abraham e Nath, 2000].

O desempenho de um sistema difuso depende criticamente da definição de suas funções de pertinência e de seu conjunto de regras. A definição destas funções normalmente é feita por um especialista, que utiliza julgamentos subjetivos e particulares na especificação das formas e valores das funções. A sintonia desses parâmetros é realizada usualmente de forma heurística, o que torna o processo muito lento e subjetivo. A utilização das redes neurais artificiais para gerar os valores das funções de pertinência tem obtido ótimos resultados, pois reduz drasticamente o tempo de configuração do sistema [Jang, 1993].

Nas próximas seções são apresentadas duas abordagens neuro-difusas, investigadas com o objetivo de integrarem o sistema de reconhecimento de padrões de um nariz artificial. Elas compreendem a rede *Feature-weighted Detector* e a rede neural evolutiva *Evolving Fuzzy Neural Network*.

4.2.1 Rede Feature-weighted Detector

A abordagem neuro-difusa *Feature-weighted Detector* [Li et al., 2002], consiste de uma rede com conexões de memória para classificação de padrões e conexões de pesos para seleção de atributos. Esta abordagem possui a capacidade de selecionar as características importantes de um conjunto de dados, enquanto mantém a taxa máxima de reconhecimento de padrões. Além da capacidade de selecionar atributos e classificar padrões, o método possibilita a extração do conhecimento adquirido pela rede no

processo de aprendizagem na forma de um conjunto de regras interpretáveis do tipo *Se-Então*.

A rede Feature-weighted Detector, como mostra a Figura 4.4, consiste de uma arquitetura com quatro camadas: camada de Entrada (E), camada de Casamento (C), camada de Detecção (D) e camada de Saída (S). Abaixo é apresentada uma descrição detalhada da rede.

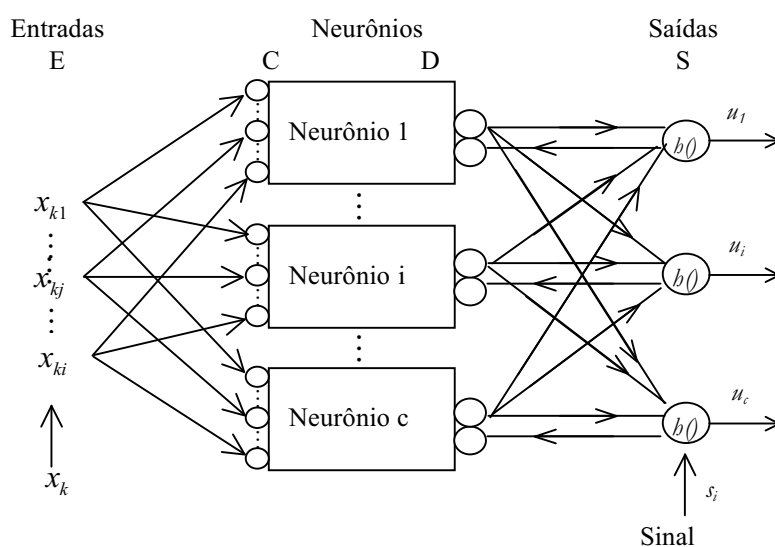


Figura 4.4 - Modelo da rede *Feature-weighted Detector*

4.2.1.1 Dinâmica de Funcionamento

Para facilitar a compreensão do modelo, será considerado o funcionamento de uma unidade de processamento (neurônio) isoladamente, como mostra a Figura 4.5. A camada *E* é responsável por fornecer os padrões de entrada à rede. Cada unidade de processamento recebe todas as entradas apresentadas à rede.

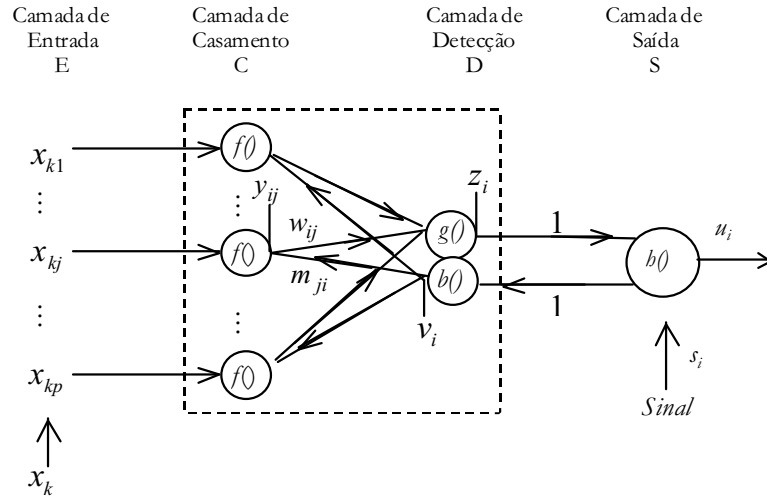


Figura 4.5 - Comportamento estrutural e de interconexão do neurônio

A camada C efetua o casamento dos padrões de entrada com as conexões de memória $\{m_{ji}\}$, que são responsáveis pela classificação difusa dos dados. Para isto, a camada recebe entradas através de conexões com as camadas E e D . A função de ativação dos nodos desta camada é uma função comparativa $f(\cdot)$ descrita pela Equação (4.1).

$$y_{ij} = (x_{kj} - m_{ji}), \quad j = 1, 2, \dots, p; \quad i = 1, 2, \dots, c; \quad k = 1, 2, \dots, N \quad (4.1)$$

Na Equação (4.1), p representa o número de atributos, c o número de classes e N o número de padrões.

De acordo com o fluxo de processamento, a camada D possui dois tipos de unidades de processamento: *propagação* e *retropropagação*. As unidades de propagação recebem p entradas através das conexões $\{w_{ij}\}$ com a camada C e fornecem dados para a camada S , a partir de conexões de pesos fixos, com valor 1. A função de ativação dos nodos da camada D é uma função gaussiana $g(\cdot)$, descrita pela Equação (4.2).

$$z_j = \exp\left(-\frac{1}{2\sigma^2} \cdot \sum_{j=1}^p w_{ij}^2 \cdot y_{ij}^2\right) \quad (4.2)$$

As unidades de retropropagação recebem entradas das conexões vindas da camada S , também com pesos fixos em 1. A função de ativação $b(.)$ destas unidades é uma função identidade.

A camada S é responsável por fornecer a classificação da rede. Esta camada funciona como uma função de pertinência difusa, efetuando a normalização de todos os valores provenientes da camada D . A camada S recebe $c+1$ entradas, c entradas através das conexões com a camada D e uma entrada chamada de sinal. A função de ativação da camada S , $h(.)$ é descrita pela Equação (4.3).

$$u_i = z_i / \sum_{j=1}^c z_j \quad (4.3)$$

4.2.1.2 Aprendizagem

A rede *Feature-weighted Detector* utiliza aprendizagem supervisionada e não-supervisionada em seu treinamento. A inicialização dos valores para as conexões de memória é baseada nos centros dos dados de treinamento. Esta inicialização é realizada calculando-se o centróide de cada uma das variáveis em relação aos exemplos de treinamento. As conexões de pesos são inicializadas com valores fixos em 1.

O ajuste das conexões de memória $\{m_i\}$, responsáveis pela classificação difusa dos padrões, é realizado de modo não-supervisionado. Esta atualização é descrita pela Equação (4.4).

$$\Delta m_i = \alpha \cdot (1 - t/T) \cdot u_i(x_k) \cdot (x_k - m_i) \quad (4.4)$$

onde

$$m_i(n+1) = m_i(n) + \Delta m_i(n)$$

Nesta equação, x_k representa a k -ésima entrada e $\alpha \in [0,1]$ se refere à taxa de aprendizado temporal cujo valor é definido pelo usuário. T corresponde ao número de épocas de treinamento e t indica o passo de treinamento. E u_i é a saída atual da rede.

O ajuste das conexões de pesos $\{w_{ij}\}$, responsáveis pela seleção dos atributos mais importantes para a classificação, é realizado de modo supervisionado de acordo com a Equação (4.5).

$$\Delta w_{ij} = \frac{\beta}{\sigma^2 \cdot s^2} \cdot (u_i(x_k) - d_i) \cdot (s - z_i) \cdot w_{ij} \cdot z_i \cdot y_{ij}^2 \quad (4.5)$$

onde

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n)$$

Na Equação (4.5), $s = \sum_{i=1}^c z_i$, $\beta > 0$ é a taxa de aprendizado, $\sigma > 0$ representa a nebulosidade no agrupamento, z_i é a saída da camada de detecção e d_i é o valor da saída desejada para a unidade de processamento i . Para facilitar a compreensão, $0 \leq w_{ij} \leq 1$.

O erro de treinamento da fase supervisionada é dada pela Equação (4.6).

$$E = \frac{1}{2} \cdot \sum_{k=1}^N \sum_{i=1}^c (u_i(x_k) - d_i)^2 \quad (4.6)$$

Abaixo são descritos os passos que contituem o algoritmo de aprendizado da rede *Feature-weighted Detector*.

1. Escolher os parâmetros de inicialização da: nebulosidade no agrupamento com $\sigma > 0$, taxa de aprendizado temporal com $\alpha \in [0,1]$, taxa de aprendizado $\beta > 0$, Erro mínimo de treinamento $\varepsilon > 0$ e o número máximo de iterações T .
2. Inicialize as conexões de memória $\{m_i(0)\}$, usando c exemplos escolhidos randomicamente de $\{x_k\} (k = 1, 2, \dots, N)$ e $w_{ij}(0) = 1$ para cada i e j .
3. PARA $t = 1, 2, \dots, T$; Para $k = 1, 2, \dots, N$ FAÇA
 - a. Calcule a saída da rede $\{u_i\}$ usando a Equação (4.3).
 - b. Atualize as conexões de memória $\{m_i(t)\}$ usando a Equação (4.4).
 - c. Atualize as conexões de pesos $\{w_{ij}(t)\}$ usando a Equação (4.5).
 - d. Próximo k .

4. Calcule o erro de treinamento E usando a Equação (4.6).
5. SE o erro de treinamento $E < \varepsilon$, Ou SE o número de iterações $t > T$
PARE, SENÃO próximo t .

4.2.1.3 Seleção de Atributos e Extração de Regras

A rede *Feature-weighted Detector* possibilita que o conhecimento adquirido pela rede durante a fase de aprendizado seja representado por regras explicativas difusas do tipo *Se-Então*. O processo de extração de regras da rede é realizado através das conexões de memória $\{m_{ji}\}$, obtidas durante a fase de treinamento. Este processo utiliza a Equação (4.7), derivada das Equações 4.1 e 4.2.

$$z_i = \exp\left[-\frac{1}{2\sigma^2} \cdot w_{i1}^2 \cdot (x_1 - m_{1i})^2\right] \cdot \exp\left[-\frac{1}{2\sigma^2} \cdot w_{i2}^2 \cdot (x_2 - m_{2i})^2\right] \cdot \dots \cdot \exp\left[-\frac{1}{2\sigma^2} \cdot w_{in}^2 \cdot (x_n - m_{ni})^2\right] \quad (4.7)$$

$$= U_{A_{i1}}(x_1) \cdot U_{A_{i2}}(x_2) \cdot \dots \cdot U_{A_{in}}(x_n), \quad i = 1, 2, \dots, c \quad n = 1, 2, \dots, p$$

Os valores de z_i correspondem às saídas da rede para cada uma das classes. O conjunto difuso do atributo n da classe i é representado por A_{in} e $U_{A_{in}}$ é a função de pertinência de A_{in} .

Cada saída da rede é normalizada utilizando a Equação (4.8), a fim de que o grau de pertinência do padrão x para a classe c_i mantenha-se no intervalo de valores [0 e 1] e a soma das saídas dos neurônios seja igual a 1.

$$u_{ci} = U_{A_{i1}}(x_1) \cdot U_{A_{i2}}(x_2) \cdot \dots \cdot U_{A_{in}}(x_n) / \sum_{j=1}^c U_{A_{j1}}(x_1) \cdot U_{A_{j2}}(x_2) \cdot \dots \cdot U_{A_{jn}}(x_n), \quad i = 1, 2, \dots, n \quad (4.8)$$

A construção das regras difusas é realizada da seguinte forma:

Se x_1 é aproximadamente m_{1i} **E** ... **E** x_n é aproximadamente m_{ni} $i = 1, 2, \dots, c$
Então x pertence à classe c_i $n = 1, 2, \dots, p$

O antecedente da regra corresponde ao lado direito da Equação (4.7) e o conseqüente ao lado esquerdo. Os números difusos *aproximadamente* m_{ji} são representados pelos conjuntos difusos $A_{i1}, \dots, A_{in}$ da Equação (4.7).

Com relação à seleção de atributos, esta é realizada a partir das conexões de pesos $\{w_{ij}\}$, cujos valores representam o grau de contribuição do atributo j para o agrupamento i . Por exemplo, $w_{ij} = 0$ indica que o atributo j não tem contribuição alguma na formação do agrupamento i e poderia ser retirado da base sem prejuízo para a classificação. Os atributos com menor relevância na formação dos agrupamentos podem ser excluídos da base de dados, sem prejuízo algum no desempenho do classificador.

4.2.2 Rede Evolving Fuzzy Neural Network

A abordagem evolutiva neuro-difusa *Evolving Fuzzy Neural Network* [Kasabov, 1998b], é uma estrutura conexionista que possui características como: aprendizado rápido (possivelmente em modo *one-pass learning*), incremental e *on-line*; sua estrutura conexionista pode ser interpretada por regras difusas; utiliza regras de classificação para o crescimento, poda e agregação de sua topologia; e suas unidades de processamento e as conexões entre estas unidades são criadas, eliminadas e agregadas dinamicamente durante a operação da rede.

Toda a estrutura da rede Evolving Fuzzy Neural Network é construída sobre um conjunto de regras difusas, onde é possível modificar dinamicamente a topologia da rede. Assim, o modelo pode se adaptar a novos dados de treinamento (inclusive inserindo novas classes), de uma maneira *on-line* e aprender incrementalmente a partir dos exemplos de treinamento. Como a estrutura conexionista da rede pode ser interpretada por regras difusas, todo o conhecimento adquirido pela rede pode ser expresso na forma de regras explicativas difusas do tipo *Se-Então*.

Desta forma, uma Evolving Fuzzy Neural Network é um sistema conexionista que aprende a partir de exemplos, manipula o conhecimento a partir de regras difusas e realiza inserção, extração e agregação destas regras na arquitetura da rede. As redes

Evolving Fuzzy Neural Networks operam em um modo *on-line* e aprendem incrementalmente, através de características locais dos dados.

A arquitetura Evolving Fuzzy Neural Network é denominada evolutiva porque sua topologia cresce com a apresentação de novos dados na fase de aprendizado e regularmente diminui através de poda (*prunning*), ou através da agregação de unidades de processamento de regras.

4.2.2.1 Dinâmica de funcionamento

As redes Evolving Fuzzy Neural Networks constroem um conjunto de regras difusas a partir dos exemplos de treinamento e realizam inferência difusa de um modo conexionista [Jang, 1993]. Este modelo se desenvolve (ou evolui) de acordo com os princípios *ECOS* (*Evolving Connectionist Systems*) apresentados por Kasabov (1998).

A rede Evolving Fuzzy Neural Network adapta mecanismos de várias técnicas de aprendizagem [Kasabov, 1999]. Porém, em seu funcionamento todas as unidades de processamento são criadas dinamicamente, durante a fase de aprendizado da rede (possivelmente em modo *one-pass learning*). Cada uma das unidades de processamento da rede representa uma ou mais funções de pertinência difusa, as quais podem ser modificadas durante o treinamento da rede.

A arquitetura Evolving Fuzzy Neural Network possui cinco camadas, como mostra a Figura 4.6. A camada de entrada recebe o vetor de entrada (x, y) , correspondente aos atributos de entrada e a saída desejada de cada exemplo, respectivamente. Cada atributo de entrada é representado por um grupo de unidades de processamento que realizam a quantificação difusa destas variáveis.

Na segunda camada da rede as unidades de processamento são espacialmente arranjadas para representar as entradas em um espaço difuso, denominado espaço de entrada do problema. Diferentes funções de pertinência podem ser agregadas a cada uma das unidades de processamento (por exemplo, funções triangulares, gaussianas, etc.).

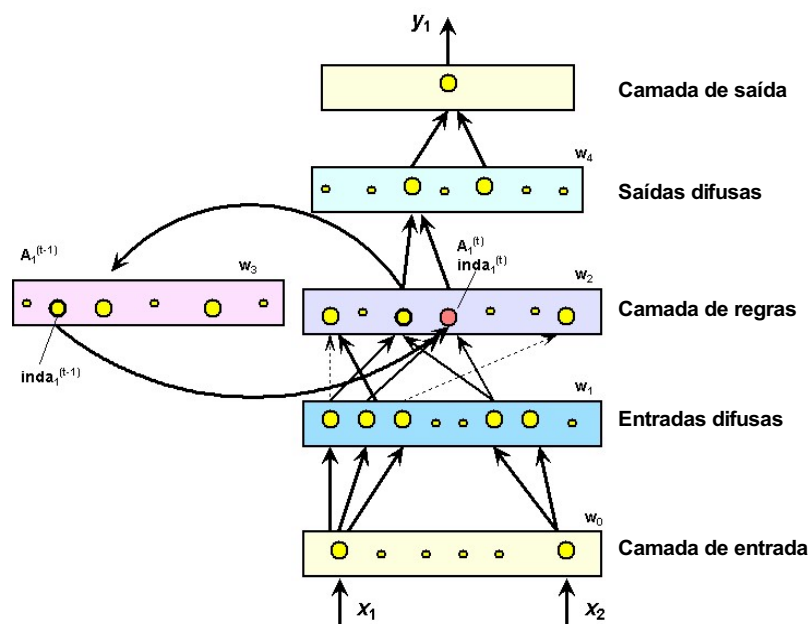


Figura 4.6 - Arquitetura da rede Evolving Fuzzy Neural Network (adaptado de Kasabov (2001))

Novas unidades de processamento podem ser inseridas dinamicamente na segunda camada durante a adaptação da rede. Isso ocorre quando, para um dado vetor de entrada, uma variável não pode ser representada por nenhuma das unidades existentes. A correspondência entre uma nova variável e as funções de pertinência existentes nas unidades de processamento são medidas pela similaridade entre elas. Por exemplo, uma nova unidade de processamento difuso, ou uma nova unidade de entrada podem ser criadas durante a fase de adaptação da rede se a dissimilaridade entre a variável de entrada e os nodos de processamento ultrapassar um limiar pré-estabelecido pelo usuário.

A terceira camada do modelo contém as unidades de processamento correspondentes às regras difusas, estes nodos representam a associação dos padrões de entrada e saída, e são adaptados através de um aprendizado híbrido: supervisionado e não-supervisionado. Cada unidade de processamento ou regra representa um protótipo da associação entre exemplos de entrada e saída, graficamente representado como uma associação entre hiper-esferas de espaços de entrada e saída difusos do problema.

A Figura 4.7 apresenta uma unidade de processamento de regras que representa uma associação entre duas hiper-esferas (entrada e saída). Espacialmente, o nodo de regras se move no espaço quando um novo par de entradas difusas (x_f, y_f) é aprendido.

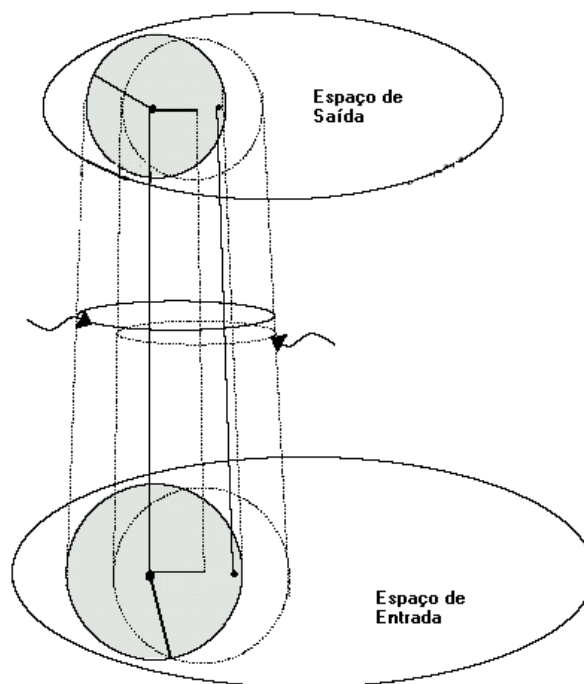


Figura 4.7 - Associação entre hiper-esferas de entrada e saída

Durante a fase de aprendizado uma camada de memória de curta-duração pode ser utilizada para adaptação da rede, através de conexões de *feedback* provenientes da camada de regras, como mostra a Figura 4.6. Como a rede utiliza aprendizado local, nodos de regras podem ser modificados ou excluídos sem prejudicar o aprendizado global do sistema ou o conhecimento previamente adquirido pela rede.

Cada nodo de regras r é definido por dois vetores de conexões $W1(r)$ e $W2(r)$. Onde a conexão $W1(r)$ é ajustada com aprendizagem supervisionada, baseado no erro de saída. E a conexão $W2(r)$ é ajustada de forma não-supervisionada, baseado na medida de similaridade entre características locais do espaço de entrada e as funções de pertinência existentes.

A quarta camada de unidades de processamento representa a quantificação difusa das variáveis de saída, no espaço de saída do problema. Esta representação é similar à representação difusa das unidades de processamento de entrada.

A quinta camada representa os valores reais das variáveis de saída. A saída é um valor real que indica a grau de confiança associado a cada classe.

O processo de evolução (aprendizagem) da rede pode ser baseado em qualquer uma das seguintes suposições: (1) nenhum nodo de regras existe antes do início da fase de aprendizagem, e todos são criados (gerados) durante o processo de evolução da rede; ou (2) existe um conjunto inicial de regras com conhecimento prévio sobre o problema, onde as conexões entre os exemplos de entrada e saída são definidas e atualizadas a partir deste conjunto de regras, durante a evolução da rede.

O último caso é o mais biologicamente plausível. Da mesma forma que, por exemplo, uma criança ao nascer possui todos os neurônios e conexões sinápticas em seu cérebro, mas suas estruturas cognitivas e funcionais evoluem de acordo com seu desenvolvimento pessoal, a rede Evolving Fuzzy Neural Network adapta seu conhecimento a partir dos nodos de regras já existentes e cria novas conexões ou novos nodos durante o aprendizado. O algoritmo de treinamento apresentado na próxima seção não faz distinção entre estas duas abordagens.

Na camada de regras cada nodo de regras difusas, por exemplo, r_j , representa uma associação entre hiper-esferas do espaço de entrada e saída difusas do problema, como mostra a Figura 4.7. Os valores da conexão $W1(r_j)$ representam as coordenadas do centro da esfera no espaço de entrada difuso, e $W2(r_j)$ representa as coordenadas no espaço de saída difuso. O raio da hiper-esfera de entrada da unidade de processamento r_j é definido como $R_j = 1 - S_j$. Onde S_j é o parâmetro de limiar de sensibilidade *sthr*, que define a ativação mínima no nodo de regras r_j obtido por um novo vetor x de entrada em um novo exemplo (x, y) . O grau desta ativação determina se este exemplo é associado a uma regra existente ou uma nova regra é criada.

O par difuso de entrada-saída (x_j, y_j) será alocado ao nodo de regras r_j se x_j estiver no campo receptivo de entrada r_j (dentro do raio da hiper-esfera de entrada), e se

y_f estiver na hiper-esfera receptiva de saída r_j . Isto é garantido através de duas condições: (1) se a diferença difusa local entre x_f e $WI(r_j)$ for menor que o raio r_j ; e (2) se o erro de saída normalizado, descrito pela Equação (4.9), for menor que o limiar de erro E .

$$Err = ||y - y'|| / N_{out} \quad (4.9)$$

Onde N_{out} é o número de saídas da rede, y e y' são os vetores da saída desejado e obtido pela rede, respectivamente. O parâmetro de limiar de erro E ajusta a tolerância de erro do sistema.

4.2.2.2 Inserção, Adaptação, Agregação e Extração de regras

Sendo o sistema Evolving Fuzzy Neural Network baseado em regras, a manipulação adaptativa destas regras é essencial para seu funcionamento. Por manipulação se entende a inserção, extração, agregação e adaptação destas regras na topologia da rede.

Em qualquer fase do processo de aprendizagem (evolução), regras difusas ou exatas (*crisp*) podem ser introduzidas e extraídas de uma estrutura Evolving Fuzzy Neural Network. A inserção de regras difusas é obtida a partir do ajuste de uma nova unidade de processamento de regras r_j , para cada nova regra criada. Por exemplo:

A regra difusa: *SE x_1 for Pequeno e x_2 for Pequeno ENTÃO y é Pequeno*; pode ser introduzida na rede ajustando com valor 1 as conexões de um novo nodo de regras com os nodos de entrada x_1 -Pequeno e x_2 -Pequeno e saída y -Pequeno. O restante das conexões permanecem com valor igual a zero.

Da mesma forma, uma regra exata: *SE x_1 for 3.4 e x_2 for 6.7 ENTÃO y é 9.5*; pode ser introduzida na rede evolutiva. Neste caso, os graus de pertinência a que os valores de entrada $x_1=3.4$ e $x_2=6.7$ e o valor de saída $y=9.5$ pertencem, são calculados e inseridos junto a seus valores correspondentes na conexão (nodo) de regras criada.

O processo de extração de regras funciona da seguinte forma, cada unidade de processamento de regras r_j pode ser expressa como uma regra difusa do tipo *Se-Então*, como no exemplo abaixo.

r_j ; SE x_1 for 0.85 **Pequeno** e x_1 for 0.15 **Médio** e x_2 for 0.7 **Pequeno** e x_2 for 0.3 **Médio** (com Raio do campo receptivo $R_j=0.1$ e $\max Radius_j=0.2$) ENTÃO y é 0.2 **Pequeno** e y é 0.8 **Grande**; (o número de exemplos associados a esta regra é de 20 exemplos em 175).

Nesta regra os números junto aos rótulos difusos (Pequeno, Médio e Grande) denotam os graus de pertinência pertencentes aos centros das hiper-esferas de entrada e saída. Somente valores de pertinência maiores que limiares $T1$ (entrada) e $T2$ (saída), configurados pelo usuário, são descritos por regras.

Outra técnica baseada em conhecimento aplicada às Evolving Fuzzy Neural Networks é a agregação de unidades de processamento de regras. Com esta técnica, diversas unidades de processamento próximas podem ser fundidas em uma unidade de processamento comum. A idéia básica deste processo é ilustrada na Figura 4.8.

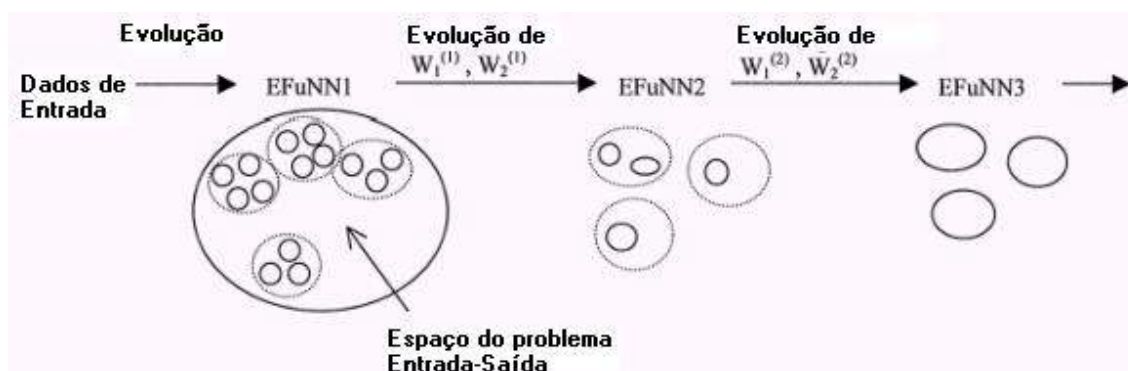


Figura 4.8 – Processo de agregação e otimização da arquitetura (adaptado de Kasabov (2001))

O processo de agregação e otimização da estrutura Evolving Fuzzy Neural Network pode ser visto como um processo de abstração de conhecimento [Kasabov, 2001 p. 15]. Para simplificar, consideremos um exemplo com três nodos de regras r_1 , r_2 , e r_3 , como apresentado na Figura 4.9. O processo de agregação é realizado a partir de regras de agregação. Nas Equações 4.10 e 4.11 são apresentadas as regras de agregação utilizadas na rede Evolving Fuzzy Neural Network, qualquer uma das duas equações pode ser utilizada para calcular os valores das conexões $W1$ e $W2$ de um novo nodo de regras agregado r_{agg} . As regras descritas nas Equações 4.10 e 4.11 são baseadas nos centros geométricos e pesos estatisticamente centrados, respectivamente.

- Como o centro geométrico de três nodos:

$$W1(r_{agg}) = (W1(r_1) + W1(r_2) + W1(r_3)) / 3 \quad (4.10)$$

- Como um peso estatisticamente centrado:

$$W2(r_{agg}) = (W2(r_1) N_{ex}(r_1) + W2(r_2) N_{ex}(r_2) + W2(r_3) N_{ex}(r_3)) / N_{sum} \quad (4.11)$$

onde $N_{ex}(r_{agg}) = N_{sum} = N_{ex}(r_1) + N_{ex}(r_2) + N_{ex}(r_3)$; $R_{ragg} = d(W1(r_{agg}), W1(r_j)) + R_j = R_{max}$, e $d(.)$ corresponde a distância entre as regras no espaço; r_j é o nodo de regras entre os três nodos que tem a distância máxima do novo nodo de regras r_{agg} ; o parâmetro R_j é o raio do campo receptivo atual. Os três nodos de regras somente serão agregados se o raio do nodo de agregação r_{agg} for menor que um raio máximo predefinido como R_{max} .

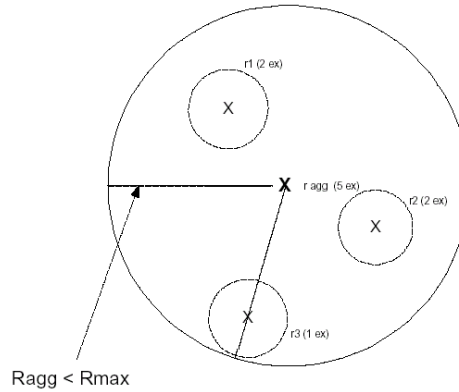


Figura 4.9 – Agregação de nodos de regras em uma rede Evolving Fuzzy Neural Network

Através da criação de novos nodos e sua consecutiva agregação, um sistema Evolving Fuzzy Neural Network pode se ajustar a mudanças que venham a ocorrer nos dados ou processo representado com o passar do tempo, preservando suas potencialidades de generalização.

Na seqüência é apresentado um exemplo das regras difusas do tipo *Se-Então* que podem ser extraídas da rede Evolving Fuzzy Neural Network:

Rule 1:

if [1] (1 0.312) (2 0.688)
 [2] (1 0.611) (2 0.389)
 [3] (1 0.608) (2 0.392)
 [4] (1 0.954)
 [5] (1 0.586) (2 0.414)
 [6] (1 0.526) (2 0.474)
then (1 0.727) (2 0.273)

Rule 2:

if [1] (2 0.924)
 [2] (2 0.901)
 [3] (1 0.118) (2 0.882)
 [4] (1 0.954)
 [5] (1 0.121) (2 0.879)
 [6] (1 0.135) (2 0.865)
then (1 0.857)

Rule 3:

if [1] (1 0.429) (2 0.571)
 [2] (1 0.782) (2 0.218)
 [3] (1 0.817) (2 0.183)
 [4] (1 0.955)
 [5] (1 0.773) (2 0.227)
 [6] (1 0.883) (2 0.117)
then (1 0.725) (2 0.272)

Na premissa da regra, os valores entre colchetes [.] representam os atributos ou variáveis das instâncias de entrada. Os valores entre parênteses (.) representam respectivamente a função de pertinência (Pequena, Média ou Grande) e seu grau associado para a referida variável. Da mesma forma na conclusão da regra, a representação correspondente à função de pertinência e seu valor associado.

4.2.2.3 Aprendizagem

O algoritmo de treinamento da rede Evolving Fuzzy Neural Network [Kasabov, 1998b] é baseado no princípio de que nodos de processamento somente existem se forem realmente necessários. Enquanto cada exemplo de treinamento é apresentado, os valores de ativação das unidades de processamento da camada de regras e o erro sobre as unidades de processamento de saída são examinados. Se a ativação máxima nos nodos de regras estiver abaixo de um limiar estabelecido pelo usuário (limiar de sensibilidade) o nodo de regras é suprimido. Se o erro do nodo de saída estiver acima do valor de limiar também determinado pelo usuário (limiar de erro), um novo nodo de regras é

adicionado. E finalmente, se o nodo de saída mais ativado não for o vencedor desejado, uma nova unidade de saída é então adicionada.

O presente algoritmo é baseado na combinação das similaridades difusas entre os exemplos dos dados que são apresentados ao modelo, o número e à idade (tempo em que os nodos de regras existem) dos nodos de regras existentes, e a frequência de ativação destes nodos.

Abaixo são descritos os passos que constituem o algoritmo de aprendizado (evolução) da rede neural evolutiva Evolving Fuzzy Neural Network.

1. Criar uma nova estrutura com um número máximo de nodos e conexões com valor zero. As conexões iniciais podem ser ajustadas com a introdução de regras previamente existentes, de modo que informações prévias possam ser usadas antes do início do processo de aprendizagem da rede (evolução). Se inicialmente não houver nenhum nodo de regras conectando os nodos de entrada e saída, *conectar* o primeiro nodo $rn=1$ para representar o primeiro exemplo $EX=x_1$ e ajustar o peso da conexão de entrada $W1(rn)$ e da conexão de saída $W2(rn)$ como segue:

<Conectar um novo nodo de regras rn para representar um exemplo EX >:

$$W1(rn)=EX; W2(rn)=TE \quad (4.12)$$

onde TE é o vetor de saída difusa para o exemplo difuso EX .

2. EQUANTO <houver exemplos de treinamento> FAÇA inserir o corrente exemplo x_i , onde EX é o vetor de entrada difuso. Se existirem novas variáveis ou atributos neste exemplo, criar novos nodos de entrada e/ou saída com suas correspondentes funções de pertinência.

3. Encontrar a similaridade normalizada difusa entre o novo exemplo de treinamento EX (vetor de entrada difusa) e os padrões previamente armazenados $j=1,2,...,rn$:

$$D_j = \text{sum} (\text{abs} (EX - W1(j)) / 2) / \text{sum} (W1(j)) \quad (4.13)$$

onde sum é a soma e abs é o valor absoluto.

4. Encontrar a ativação dos nodos de regras j , para $j=1:rn$. Pode ser utilizada a *Função de Ativação de Base Radial*, ou a *Saturação Linear* nos valores de entrada D_j . Isto é:

$$A1(j) = \text{radbas} (D_j), \text{ ou } A1(j) = \text{satlin} (1-D_j) \quad (4.14)$$

onde radbas é a *Função de Base Radial* e satlin a *função de Saturação Linear*.

5. Atualizar os parâmetros locais definidos para os nodos de regras, por exemplo, idade, ativação média predefinida, etc.
6. Encontre todos os nodos j com valor da ativação $AI(j)$ acima do limiar de sensibilidade $sthr$.
7. Se não houver nenhum nodo, então *<conectar um novo nodo de regras>* usando o procedimento definido na etapa 1.

SENÃO

8. Encontrar o nodo de regras $inda1$ que possui o valor máximo de ativação ($maxa1$).
9. (a) no caso do modo de treinamento da rede ser de *um-pra-n*, propague a ativação $maxa1$ do nodo de regras $inda1$, aos nodos de saída difusa. As funções de *Saturação Linear* são usadas como funções da ativação para os nodos de saída difusos:

$$A2 = satlin(A1(inda1) * W2) \quad (4.15)$$

- (b) no caso do modo de operação ser de *muitos-pra-n*, somente os valores de ativação das unidades de processamento que estão acima do limiar de ativação $athr$ são propagados à camada seguinte.

10. Encontrar a unidade de processamento de saída vencedor $inda2$ e sua ativação $maxa2$.
11. Verificar a saída difusa desejada $indt2$ e seu valor $maxt2$.
12. Calcular o vetor de erros difusos de saída:

$$Err = A2 - TE \quad (4.16)$$

13. SE($inda2$ é diferente de $indt2$) ou ($abs(Err(inda2)) > errthr$)

<Conectar um novo nodo de regras>

onde $errthr$ é o limiar de erro.

SENÃO

14. Atualizar: (a) as conexões dos nodos de regras $k=inda1$ de entrada; e (b) as conexões de saída conforme a equação abaixo:

(a) $Dist = EX - W1(k)$; $W1(k) = W1(k) + lr1$, onde $lr1$ é a taxa de aprendizado para as conexões de entrada;

(b) $W2(k) = W2(k) + lr2$, onde $lr2$ é a taxa de aprendizado para as conexões de saída.

15. Podar os nodos de regras j e suas conexões de modo que satisfaçam à seguinte regra de poda (*prunning*):

SE (o nó (j) é VELHO) e (a ativação média $Al_{av}(j)$ é BAIXA) e (a densidade da área de vizinhança dos nodos é ELEVADA ou MODERADA) e (a soma dos novos pesos de entrada ou das novas conexões de saída é BAIXA) e (o nodo NÃO está associado com a classe correta dos nodos de saída (para tarefas da classificação somente)) ENTÃO a probabilidade do nó ser podando (j) é ELEVADA.

A regra de poda acima descrita é difusa e requer que conceitos difusos como, *VELHO*, *ELEVADO*, etc. sejam previamente definidos. Como um caso parcial, um valor fixo pode ser usado, por exemplo, um nó é velho se existir durante a evolução de uma *Evolving Fuzzy Neural Network* em mais de 60 exemplos.

16. FIM do laço SENÃO.

17. Repita etapas 2-16 para uma segunda apresentação dos mesmos dados de entrada ou para o treinamento *ECO* [Kasabov, 1988], se necessário.

4.3 Sumário do capítulo

Este capítulo objetivou apresentar as técnicas utilizadas nesta dissertação no reconhecimento dos padrões de odores. O uso de redes neurais artificiais como sistemas de reconhecimento de padrões em narizes artificiais tem sido muito investigado, a grande maioria destas abordagens faz uso da conhecida rede Perceptron Multi-Camadas.

A vantagem da utilização de redes neurais artificiais no reconhecimento de padrões é a não necessidade de conhecimento sobre os dados, devido a grande capacidade de generalização da rede a partir de dados não-lineares. Sua desvantagem é

que a definição da arquitetura ideal para a rede é baseada em um processo empírico, simples, porém exaustivo.

As abordagens investigadas neste trabalho correspondem a técnicas conexionistas e híbridas. Dentre as abordagens conexionistas, destaca-se a rede Perceptron Multi-Camadas, amplamente utilizada na classificação de odores e as Rede Neurais com Atrasos no Tempo, que fazem uso de processamento temporal. As Redes Neurais com Atrasos no Tempo levam em consideração variações dos sinais dos sensores do nariz artificial ao longo da aquisição dos dados.

As abordagens neurais híbridas investigadas correspondem às redes Feature-weighted Detector, que possibilita classificação, seleção de atributos e extração de regras da rede e a rede Evolving Fuzzy Neural Network, que possibilita aprendizado *on-line*, inserção, extração e agregação de regras difusas da topologia da rede.

No capítulo seguinte são apresentadas as técnicas de pré-processamento de dados, metodologia e critério de avaliação utilizados nos experimentos.

5

Métodos

Este capítulo apresenta as técnicas de pré-processamento de dados, metodologia e critério de avaliação utilizados nos experimentos. A seção 5.1 descreve a fase de pré-processamento dos padrões de odores. Na seção 5.2 são relacionados os aspectos observados nos experimentos. Os métodos estatísticos de comparação utilizados na análise dos resultados são descritos na seção 5.3. E a seção 5.4 é um sumário do capítulo.

5.1 Pré-processamento dos dados

5.1.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste

Em qualquer problema a ser resolvido através do uso de redes neurais artificiais com treinamento supervisionado, é necessário o uso de dados que correspondam a pares de entrada-saída já conhecidos. Muitas vezes a quantidade disponível de dados, para o treinamento do modelo conexionista, não é muito grande, mesmo assim faz-se necessário dividir a base de dados existente em dois conjuntos bem definidos: um conjunto de treinamento e um conjunto de teste.

O treinamento da rede neural artificial é realizado a partir do conjunto de dados de treinamento. É, entretanto, necessário medir a performance da rede, considerando

como ela responde a pares de entrada-saída não apresentados durante a fase de treinamento. Esta avaliação é feita a partir do conjunto de teste, que mede o grau de generalização apresentado pela rede para o problema abordado.

Obter esta informação sobre o desempenho da rede durante a fase de treinamento pode ser muito útil, pois pode evitar um problema conhecido como *overfitting* ou super-treinamento [Hayken, 1994]. Isto é feito com a utilização de um terceiro conjunto de pares de entrada-saída, retirado do conjunto de treinamento, chamado de conjunto de validação. O conjunto de validação serve para que o desempenho da rede seja medido durante a fase de treinamento e também pode ser utilizado como um mecanismo de parada precoce do treinamento [Prechelt, 1994].

Nos experimentos realizados nesta dissertação, a base de padrões de odores foi dividida em três conjuntos de dados. Os quais correspondem aos conjuntos de treinamento, validação e teste. A proporção de padrões contidos em cada um dos conjuntos foi a seguinte: 50% dos padrões de odores foram dispostos no conjunto de treinamento; 25% dos padrões correspondem ao conjunto de validação; e os 25% restantes foram utilizados no conjunto de teste. Esta divisão dos dados é sugerida pelo relatório *Proben1* [Prechelt, 1994], um conhecido *Benchmark* para o treinamento de redes neurais artificiais.

Para a obtenção dos conjuntos de treinamento, validação e teste foram criadas 3 partições de dados, baseadas no conjunto original de dados, descrito no Capítulo 3. Cada uma destas partições consiste da mesma base de dados, porém com suas instâncias dispostas em diferentes ordens. A partir destas partições, foram divididos os conjuntos de treinamento, validação e teste, descritos no parágrafo anterior. Desta forma, existem três bases de dados disponíveis para os experimentos. Estas bases contêm os mesmos valores, porém dispostos de maneira diferente e aleatória. A utilização de partições de dados é necessária para eliminar qualquer viés existente no treinamento das redes neurais artificiais [Prechelt, 1994].

5.1.2 Normalização

O pré-processamento dos dados é um processo de transformação dos dados de entrada brutos em dados com distribuição de interesse, para auxiliar na praticidade computacional, extração das características e minimização de ruídos.

Para uma maior efetividade e rapidez no treinamento de redes neurais artificiais, é aconselhável a transformação de todos os atributos de valores reais para um mesmo intervalo. Dados com atributos em escalas diferentes podem confundir o aprendizado da rede, fazendo que esta tenha maior consideração por valores com grandes magnitudes. As técnicas de normalização evitam este tipo de problema, transpondo os dados em intervalos bem definidos.

Podem ser utilizadas normalizações Euclidianas, logarítmicas ou funções trigonométricas, tais como: tangente hiperbólica, seno, cosseno, etc. Vários são os tipos de normalização existentes para o treinamento de redes neurais artificiais, eles variam de acordo com o objetivo, ou com a função de ativação utilizada nas unidades de processamento da rede neural artificial. Os intervalos mais comuns para manter os dados são de 0 a 1 ou de -1 a 1.

Nos experimentos realizados nesta dissertação, os padrões de odores foram normalizados separadamente para cada sensor do nariz artificial. Duas formas de normalização foram utilizadas:

- Normalização para que os valores ficassem dispostos entre -1 e 1.

$$x_{norm} = \frac{(2x - x_{max} - x_{min})}{x_{max} - x_{min}}, \quad (5.1)$$

- Normalização para que os valores ficassem dispostos entre 0 e 1.

$$x_{norm} = \frac{(x - x_{min})}{x_{max} - x_{min}}, \quad (5.2)$$

onde x_{norm} é o valor normalizado correspondente ao valor original x de resistência, e x_{min} e x_{max} são, respectivamente, os valores mínimo e máximo entre os 1.080 valores de resistência, correspondente ao sensor em questão.

5.1.3 Transformada de Wavelets

Considerando que os estímulos odorantes captados pelo nariz artificial são sinais, adquiridos por um sensor em um determinado espaço de tempo, parece natural que uma técnica utilizada para o processamento de sinais digitais seja uma das ferramentas indicadas para o tratamento destes padrões. A partir deste pressuposto, a *Transformada de Wavelet*¹¹ [Daubechies, 1990] foi utilizada como método de pré-processamento dos padrões de odores.

A Transformada de Wavelet é uma ferramenta matemática que possui vantagens perante técnicas clássicas de processamento de sinais, como a *Análise de Fourier* [Brighan, 1988]: suas funções bases (wavelets) não pertencem a um espaço finito de soluções, isto é, existem teoricamente infinitas possibilidades de se projetar wavelets com propriedades especiais, voltadas para aplicações específicas. A Análise de Wavelet é uma transformação contínua onde os núcleos de integração usados na transformação do sinal são denominados de wavelets. Estas wavelets são essencialmente utilizadas de duas maneiras no estudo de processos ou sinais: (1) como um núcleo de integração para analisar a extração de informações sobre um processo; e (2) como uma base de representação ou caracterização do mesmo.

O conceito da Transformada de Wavelet foi formalizado pela primeira vez na década de 80, através de uma série de artigos escritos por Morlet (1981). Entretanto, foi no início do século passado que o físico Alfred Haar (1910) introduziu um sistema completo de funções ortogonais com muitas propriedades e características interessantes. Características que hoje fazem das wavelets uma das ferramentas matemáticas com um vasto campo de aplicações nas ciências mais diversas.

Em contraste com a Transformada de Fourier, que não trata claramente sinais que possuam abruptas variações no tempo ou características transitórias, a Transformada Wavelet é bem localizada tanto no tempo quanto na frequência. O potencial do método é ilustrado pela análise de vários modelos de séries, tais como:

¹¹ Wavelet significa “pequena onda”.

harmônicos, fractais e as mesmas séries contaminadas com vários tipos de singularidades.

5.1.3.1 Conceitos básicos

Como descrito nos parágrafos anteriores, a Transformada de Wavelet é uma transformação linear que, analogamente à Transformada de Fourier, permite representar um sinal no domínio das frequências que o compõe. A única diferença entre as duas transformadas são as funções de base utilizadas na decomposição do sinal.

O motivo que torna a Transformada de Fourier indesejável, para trabalhar com os dados do nariz artificial, é que na transformação do sinal do domínio do tempo para o domínio da frequência, perde-se totalmente a informação sobre a localização temporal (ou espacial) dos dados. Quando olhamos para a Transformada de Fourier de um sinal, é impossível dizer onde um evento em particular está localizado, pois o que é obtido são apenas as frequências que compõem o sinal.

Se um sinal não se altera no tempo, ou seja, se é um sinal dito “estacionário”, esta deficiência da Análise de Fourier não tem importância alguma. Entretanto, como em nosso caso, a maioria dos sinais contém numerosas características não estacionárias ou transitórias, tais como: tendências, mudanças abruptas e o início ou final de eventos. Estas características fazem da Análise Wavelet uma promissora ferramenta no processamento de sinais de odores.

Será apresentada, inicialmente, a Transformada de Wavelet considerando um sinal contínuo $f(t)$. A generalização para sinais bidimensionais $f(x,y)$, objeto deste estudo, pode ser obtida facilmente, tendo sido definida no trabalho de Antonie et al., 1993. A transformada contínua em wavelets de f é dada pela Equação (5.3).

$$F(a,b) = \int f(t)\psi_{a,b}(t)dt \quad (5.3)$$

Nesta equação, os parâmetros a e b variam continuamente em R , sendo que as funções $\psi_{a,b}$ são denominadas wavelets e definidas na Equação (5.4).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (5.4)$$

Claramente, tais funções são derivadas a partir de dilatações e translações de uma wavelet mãe ψ , a qual deve satisfazer o critério de admissibilidade. Tal critério é apresentado na Equação (5.5).

$$C_\psi = 2\pi \int \frac{|\hat{\psi}(u)|^2}{|u|} du < \infty \quad (5.5)$$

onde $\hat{\psi}(u)$ é a Transformada de Fourier de $\psi(t)$. Se $\hat{\psi}$ é uma função contínua, então C_ψ pode ser finito somente se $\hat{\psi}(0) = 0$, isto é, $\int \psi(x) dx = 0$.

Observando a Equação (5.4), percebemos que a Transformada de Wavelet depende de dois parâmetros a e b , que correspondem, respectivamente, às informações de escala e tempo. É comum se associar o inverso da variável de escala com a frequência. Assim, o sinal unidimensional $f(t)$ é desdobrado em uma transformada 2D $F(a,b)$, permitindo uma análise *tempo-frequência* do sinal, de forma semelhante à *Transformada de Fourier Janelada* [Vliet, 2003]. O semi-plano bidimensional (a,b) é comumente denominado espaço de fase.

A Equação (5.5) pode ser reescrita como o produto interno de f , com uma função wavelet $\psi_{a,b}$, isto é, $F(a,b) = \langle f, \psi_{a,b} \rangle$. Basicamente, este produto interno irá “medir” o conteúdo do espaço de fase do sinal f em torno do ponto (a,b) neste espaço. Em resumo a decomposição de um sinal com uma determinada *função wavelet* permite o tratamento da energia contida no sinal, em função tanto da dimensão espacial (ou tempo) como da escala do wavelet (frequência). Conhecendo as características do sinal que se deseja obter é possível realizar uma análise em cada componente frequencial e determinar quais componentes contêm as informações desejadas. Utilizando apenas as componentes com estas informações pode-se reconstruir um sinal, cancelando os efeitos de componentes frequências indesejados.

Existem várias famílias de wavelets, alguns exemplos são: *Harr*, *Daubechies*, *Coiflets*, *Biortogonal*, *Symlets*, *Morlet*, *Meyer* e *Chapéu Mexicano* [Misiti, 1996]. As famílias de wavelets são geradas a partir de uma wavelet mãe, que é comprimida N vezes, gerando as N ondas que compõe esta família. Para realizar a Transformada de Wavelet, normalmente utilizam-se famílias de wavelets que definem bases ortogonais, pois desta forma é possível realizar a transformada inversa.

A obtenção da Transformada de Wavelet de uma amostra é obtida da seguinte forma: primeiro é calculado o valor de C que representa quanto um intervalo do sinal está correlacionado com uma wavelet (item *a* da Figura 5.1). Este procedimento é realizado nos intervalos seguintes até o termino do sinal (conforme indicado no item *b* da Figura 5.1). Encontra-se depois a correlação do sinal para escalas maiores de wavelet (conforme ilustrado no item *c* da Figura 5.1).

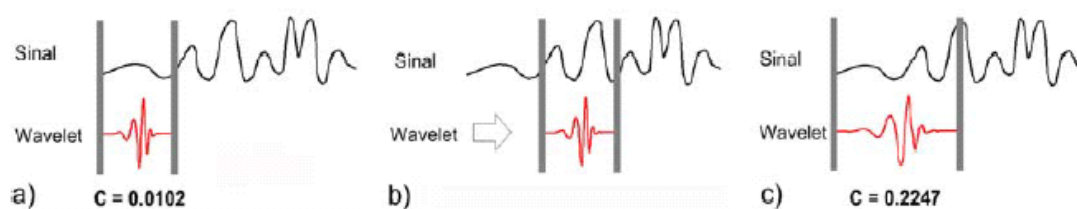


Figura 5.1 - Ilustração de como são calculados os coeficientes da Transformada de Wavelet

- a) Correlação do primeiro período do sinal com uma wavelet;
- b) Deslocamento da wavelet para obter os valores de correlação no tempo;
- c) Troca da escala da wavelet. (adaptado de [Misiti, 1996]).

Utilizando os valores de correlação do sinal (C) a cada instante com as diferentes escalas de wavelet, pode-se montar um gráfico de intensidade de correlação, onde é possível obter informação do sinal no domínio do tempo.

No entanto, digitalmente não é possível realizar estas correlações para um número infinito de escalas. O que foi solucionado a partir do desenvolvimento de um eficiente método para encontrar a Transformada Wavelet Discreta [Mallat, 1989]. O algoritmo de Mallat consiste em um banco de filtros FIR, no qual o sinal é dividido sucessivamente em sinais de alta e baixa frequência. A árvore de decomposição tradicional da transformada wavelet de um sinal s é ilustrada na Figura 5.2.

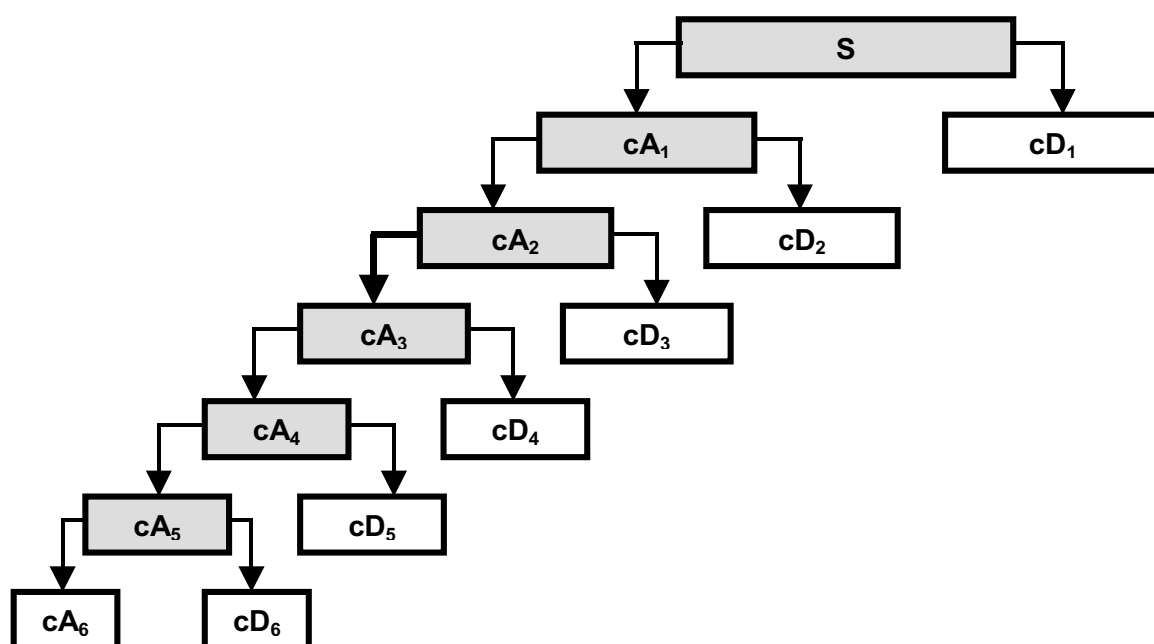


Figura 5.2 - Árvore de decomposição tradicional da transformada wavelet com seis níveis de decomposição (adaptado de [Misiti, 1996]).

Na Figura 5.2 as componentes de baixas frequências são representadas por cA_n e as de alta frequência por cD_n . A árvore de decomposição tradicional faz divisões sucessivas sempre na componente de baixa frequência. É interessante salientar que a reconstrução do sinal é realizada por meio dos coeficientes terminais de uma árvore de decomposição.

5.1.3.2 Análise de Multiresolução

Com a decomposição de um sinal utilizando a transformada wavelet é possível realizar uma análise multiresolução deste sinal, ou seja, analisando cada componente terminal da decomposição pode-se observar as características no domínio do tempo, de cada faixa de frequência contidas no sinal. Na Figura 5.3 é ilustrado um sinal (s) e os valores dos nós terminais de uma decomposição até o 3º nível, utilizando uma árvore de decomposição tradicional.

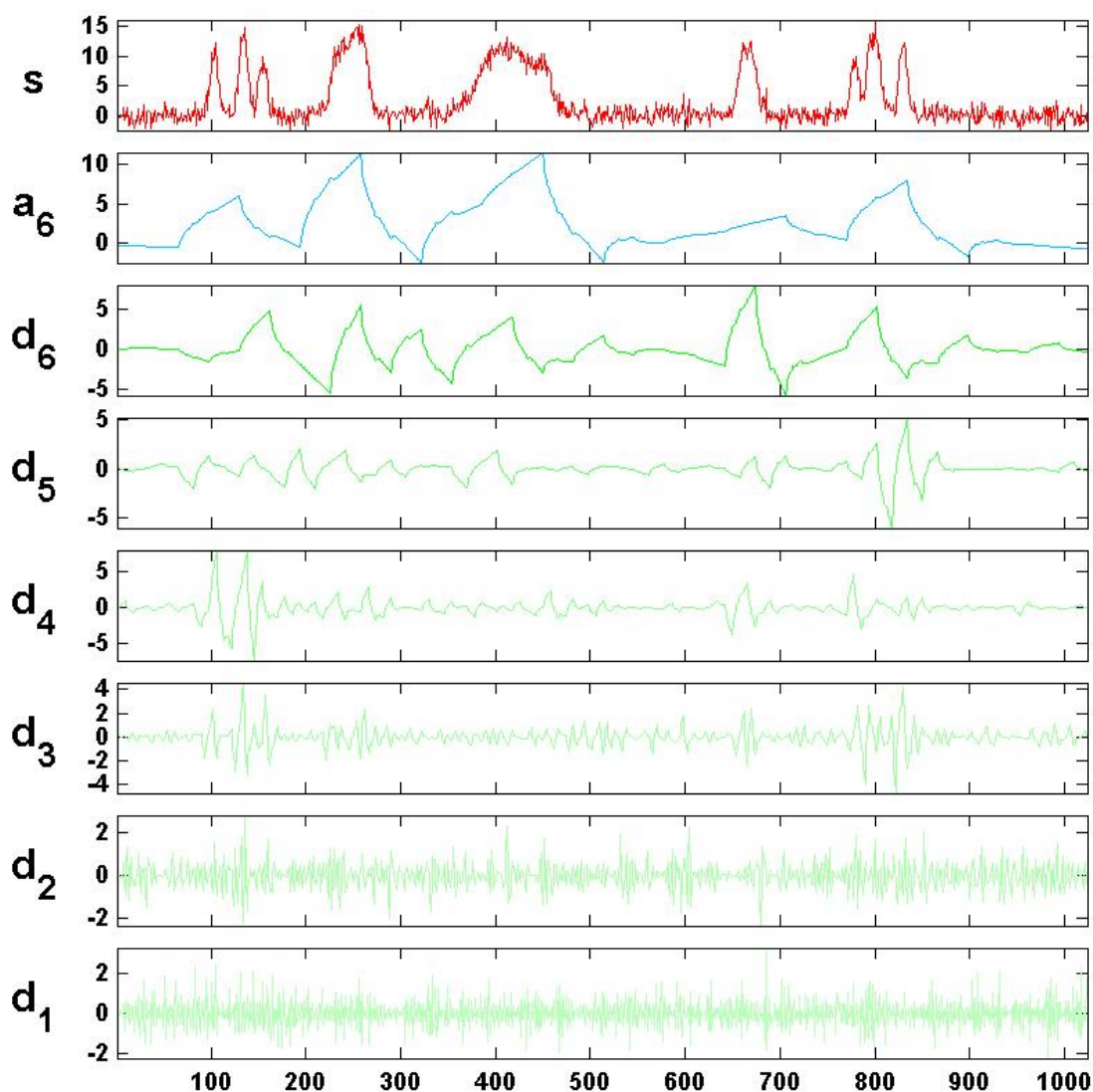


Figura 5.3 - Decomposição do sinal, s , nas componentes terminais da árvore da Figura 5.2

Pelo fato da transformada wavelet ser realizada por um banco de filtros, a seletividade das componentes frequências de um sinal em cada coeficiente (a_n e d_n) é determinada pela ordem da wavelet utilizada. Por exemplo, a *Daubechies 4* [Daubechies, 1990] seleciona melhor as frequências nos coeficientes do que a *Daubechies 2* [Daubechies, 1990] que apresenta um espalhamento maior das frequências nos coeficientes vizinhos.

Utilizando a característica multiresolução que a transformada wavelet proporciona, é possível realizar a filtragem de um sinal contaminado com algum ruído e a compressão do sinal, por meio de sucessivas decomposições deste sinal até encontrar

o ruído em algumas componentes da decomposição e então reconstruir o sinal utilizando apenas as componentes que não apresentam o ruído ou características indesejáveis.

5.1.3.3 Compressão de dados e redução de ruídos

A transformada wavelet apresenta a intensidade das componentes frequenciais que compõem um sinal. Conhecendo as características do sinal que se deseja obter é possível realizar uma análise em cada componente frequencial e determinar quais componentes contêm as informações desejadas. Utilizando apenas as componentes com estas informações pode-se reconstruir um sinal, cancelando o efeito de componentes frequenciais indesejadas.

Uma função de base também pode ser comprimida (por exemplo, período reduzido à metade, frequência crítica dobrada), no domínio do tempo (por exemplo, comprimento de onda para espectros¹², tempo para cromatogramas¹³, diferença de potencial para voltagens, etc.). Desta forma, é possível representar o sinal em escalas cada vez menores, removendo ruído sequencial em cada uma destas escalas e comprimindo o sinal. Além disso, a wavelet é transladada através do sinal, permitindo remover ruído de sinais estacionários e não estacionários.

Como descrito na Equação (5.5), a Transformada de Wavelet torna-se função dos parâmetros de escala (a e b), em que a representa a escala (compressão) e b o deslocamento. Assim sendo, cada valor de a define um tipo de filtro, isto é, um tipo de corte nas frequências.

O sinal inicial, livre de ruído, pode ser recuperado aplicando-se a *Transformada de Wavelet Inversa* relativa à wavelet mais comprimida, dada pela Equação (5.6).

$$x(t) = \int_0^{\infty} \int_{-\infty}^{\infty} w(a,b) \psi_{a,b}(t) db \frac{da}{a^2} \quad (5.6)$$

¹² Arranjo dos componentes de um feixe de energia radiante, formado quando este feixe é submetido à dispersão de modo que as ondas componentes são dispostas na ordem de seus comprimentos de onda.

¹³ Registro da concentração ou da massa dos componentes de uma amostra em função do tempo ou do volume de fase móvel, na análise das mais diversas substâncias químicas.

Através da transformada inversa, além do sinal livre de ruído, também é possível obter a representação, no domínio inicial, das características capturadas por outras wavelets, do banco de filtro. Para tanto, basta utilizar na transformada inversa a wavelet correspondente à característica desejada.

5.1.3.4 Análise dos sinais de odor

A função wavelet utilizada nesta dissertação, como filtro para eliminação do ruído e compressão dos sinais de odores, foi a transformada wavelet da família *Daubechies* de ordem 3 [Daubechies, 1988]. Este filtro wavelet foi implementado utilizando o *Toolbox Wavelets* [Misiti, 1996], do software de simulação *Matlab*¹⁴.

O sinal odorante com ruído, representado por s na Figura 5.4, foi separado nas componentes a_m (componentes de baixa frequência) e d_m (de alta frequência). A decomposição do sinal s pode ser descrita pela seguinte fórmula:

$$s = s + n = a_m + \sum_{i=1}^m d_i \quad (5.7)$$

sendo m o maior nível de decomposição, s o sinal odorante e n o ruído.

¹⁴ MATLAB, the Language of Technical Computing. Marca registrada da empresa The MathWorks, Inc.

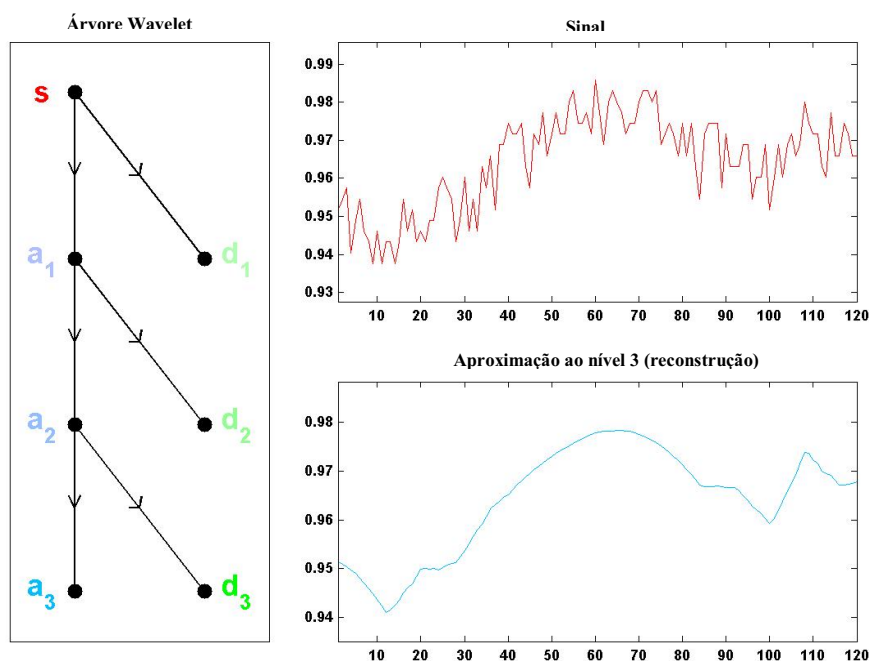


Figura 5.4 – Árvore de decomposição utilizada para filtrar o ruído e comprimir o sinal.

A partir das características do sinal de odor e do ruído, foi realizada inicialmente uma análise nas duas primeiras componentes do sinal (a_1 e d_1), observando a presença do ruído e do sinal de odor em cada uma delas. Constatou-se, conforme esperado, que a componente a_1 apresentava o ruído e uma parcela do sinal odorante. Foi então realizada uma nova decomposição nesta componente e analisadas as componentes resultantes (a_2 e d_2). Este procedimento foi feito sucessivamente até encontrar uma componente que fosse aproximadamente igual ao ruído.

Como o método de aquisição dos dados no nariz artificial induzem ruídos de baixa frequência, estes sinais foram encontrados na componente de mais baixa frequência de uma árvore de 3 níveis de decomposição (a_3). Tal análise pode ser verificada nas Figuras 5.5 e 5.6.

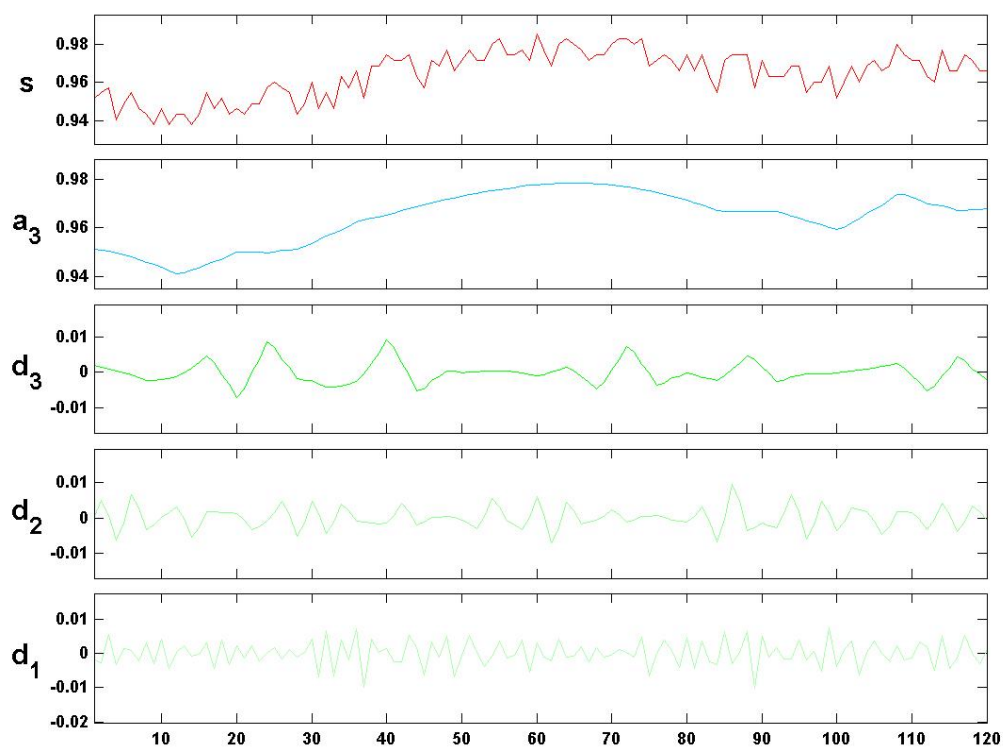


Figura 5.5 - Decomposição até o nível 3: $s = a_3 + d_3 + d_2 + d_1$.

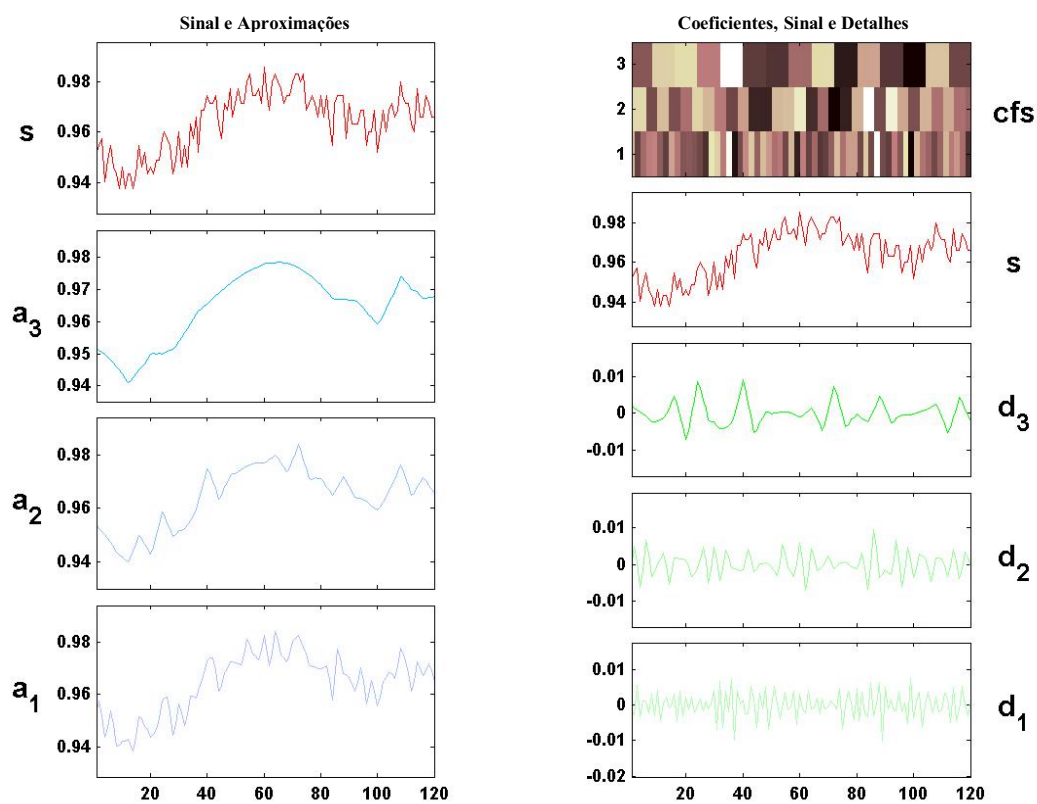


Figura 5.6 - Transformada Wavelet: Coeficientes, Sinais e Aproximações.

5.2 Aspectos observados

A medida de erro utilizada na análise dos resultados é porcentagem de erro quadrático [Prechelt, 1994], apresentada na Equação (5.8).

$$E = 100 \frac{o_{\max} - o_{\min}}{N \cdot P} \sum_{p=1}^P \sum_{i=1}^N (o_{pi} - t_{pi})^2 \quad (5.8)$$

onde, $o_{\min}$ e $o_{\max}$ são os valores de mínimo e máximo dos coeficientes de saída, na representação do problema. N é o número de unidades de saída da rede, P o número de padrões (exemplos) da base de dados considerada e t é a saída desejada da rede.

O principal parâmetro de comparação entre as abordagens investigadas foi o *Erro de Classificação* do conjunto de teste, que corresponde à razão da quantidade de padrões classificados de forma incorreta pela quantidade total de padrões.

Desta forma, os aspectos observados ao final do treinamento e que definiram o desempenho das abordagens estudadas foram os seguintes:

- ❑ Porcentagem de erro quadrático do conjunto de treinamento;
- ❑ Porcentagem de erro quadrático do conjunto de validação;
- ❑ Porcentagem de erro quadrático do conjunto de teste
- ❑ Erro de classificação do conjunto de teste;
- ❑ Número de iterações efetuadas pelo algoritmo de treinamento.

5.3 Métodos de comparação entre as abordagens

Na maioria das áreas de pesquisa há uma busca contínua por novos métodos que melhorem em algum sentido o desempenho de abordagens existentes. Na agricultura, por exemplo, buscam-se variedades mais adequadas e mais produtivas de cereais. Na computação, buscam-se técnicas que melhorem a qualidade e o desempenho dos

softwares existentes. Nestas situações é preciso comparar as técnicas usuais com os métodos alternativos propostos. Esta comparação é uma questão importante que surge freqüentemente no trabalho de pesquisa. Para se chegar a qualquer conclusão é necessário coletar informações e fazer inferências a partir de evidências experimentais ou observacionais.

Consideremos a avaliação de duas terapias na pesquisa médica, por exemplo. Se todos os portadores de uma determinada doença se comportassem de maneira idêntica em relação aos tratamentos utilizados, bastaria examinar o comportamento de um máximo de dois pacientes. A decisão sobre qual é a melhor terapia seria óbvia, e portanto, nenhuma análise seria necessária. Mas esta situação não acontece, a reação a um tratamento varia de indivíduo para indivíduo e não há um tratamento ótimo para todos os pacientes. Como, em geral, não se conhece previamente as reações de cada indivíduo sobre um determinado tratamento, prescreve-se o tratamento que apresenta em média os melhores resultados.

A generalização dos resultados obtidos no grupo estudado para uma população de interesse exige cuidados especiais no planejamento dos experimentos, sendo que fatores que afetam a resposta que está sendo avaliada devem ser controlados e considerados. O procedimento utilizado para determinar se o resultado de uma análise aponta para sim ou para não é chamado *Teste de Hipótese* [Duda et al., 2001]. Este procedimento é usado amplamente nas áreas em que as variáveis envolvidas estão sujeitas à variabilidade. O teste de hipóteses utilizado na comparação dos classificadores investigados neste trabalho é o teste t de Student.

O teste t de Student é um teste de hipóteses paramétrico, sendo os seus pressupostos de que as duas variáveis envolvidas são intervalares e que ambas tem distribuição normal. Sendo tais pressupostos verdadeiros, a variável analisada resultante pode ser matematicamente transformada, de modo a formar a distribuição t de Student; em função da qual se determina se um dado resultado é ou não estatisticamente significativo.

Maiores informações sobre o Teste de Hipóteses podem ser encontradas no Apêndice B do presente trabalho.

5.4 Sumário do capítulo

Este capítulo teve como objetivo apresentar as técnicas de pré-processamento de dados utilizadas nos experimentos, correspondendo às técnicas de Normalização e Transformada de Wavelet. O capítulo descreveu ainda os aspectos observados nos experimentos e a técnica de teste de hipótese, que será utilizada na análise dos experimentos e comparação dos resultados.

O próximo capítulo apresenta o estudo de caso, com os experimentos realizados em cada um dos classificadores apresentados no Capítulo 4.

6

Experimentos

Este capítulo descreve os experimentos realizados nesta dissertação. A seção 6.1 expõe os resultados dos experimentos com a rede Perceptron Multi-Camadas. Na seção 6.2 são descritos os resultados dos experimentos realizados com a Rede Neural com Atrasos no Tempo. Os experimentos com a rede Feature-weighted Detector são expostos na seção 6.3. A seção 6.4 descreve os resultados do treinamento da rede Evolving Fuzzy Neural Network. Na seção 6.5 é apresentada uma comparação entre as abordagens utilizadas nesta dissertação. E a seção 6.6 é um sumário do capítulo.

6.1 Experimentos com a rede Perceptron Multi-Camadas

Nesta seção serão apresentados os experimentos realizados com a rede Perceptron Multi-Camadas [Rumelhart et al., 1986], na classificação dos gases provenientes da indústria petrolífera descritos no Capítulo 3.

6.1.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste

A base de dados contendo os 5.400 padrões de odores foi dividida em três partições de dados, cada uma contendo a mesma quantidade de padrões (5.400 padrões), porém com

instâncias dispostas de maneira diferente e aleatória. Sobre cada uma destas partições foram criados 3 conjuntos de dados distintos: conjunto de treinamento para a rede neural artificial, correspondendo a 50% da quantidade total de padrões; conjunto de validação, com 25% do total de padrões da base; e o conjunto de teste, correspondendo aos 25% restantes. Esta divisão dos dados é sugerida pelo relatório *Proben1* [Prechelt, 1994].

É importante ressaltar que o conjunto de treinamento possui 50% dos padrões de cada um dos gases (para cada gás, 300 padrões são escolhidos de forma aleatória entre os 600 padrões de odores existentes). O mesmo procedimento é realizado com os conjuntos de validação e teste, de acordo com suas respectivas proporções, de forma que os cinco gases sejam representados pela mesma quantidade de dados em cada um dos três conjuntos de dados.

6.1.2 Pré-processamento dos dados

Nas simulações realizadas se utilizou os dois métodos de pré-processamento descritos na seção 5.1. Desta forma, dois experimentos foram realizados: um experimento onde os dados foram normalizados de modo que ficassem compreendidos na faixa de valores entre -1 e 1 ; e um experimento que utilizou a Transformada de Wavelet como método de pré-processamento.

6.1.3 Topologia utilizada

Foi utilizada uma rede neural Perceptron Multi-Camadas, contendo apenas uma camada intermediária. A camada de entrada possui oito unidades de processamento, as quais correspondem aos valores de cada um dos oito sensores do nariz artificial. A camada de saída possui cinco unidades de processamento, cada uma correspondendo a um dos cinco gases a serem classificados. Neste experimento foi utilizada a codificação *1-of-m*, onde m é o número de classes [Prechelt, 1994].

As unidades de processamento das camadas, intermediária e de saída possuem função de ativação *tangente hiperbólica* [Prechelt, 1994]. A rede contém todas as possíveis conexões de propagação direta (*feedforward*) entre camadas adjacentes, sem possuir conexões que ligam unidades de processamento de camadas não-adjacentes. Em cada um dos experimentos, foram treinadas seis topologias de rede distintas (com 4, 8, 12, 16, 20 e 30 unidades de processamento na camada intermediária), das quais objetivou-se avaliar os resultados.

6.1.4 Metodologia de treinamento

Como descrito no capítulo anterior o algoritmo de treinamento utilizado é uma versão do método de Levenberg-Marquardt, descrita em [Lang e Hilton, 1988]. Para cada topologia, foram realizadas 30 (trinta) inicializações de valores diferentes e aleatórios para as conexões entre as camadas de processamento. Em todos os casos, a quantidade máxima de iterações escolhida foi de 5.000 iterações.

A parada do treinamento ocorre se for satisfeito o critério GL_5 do *Proben1* [Prechelt, 1994] por duas vezes (para evitar que o treinamento seja interrompido por oscilações iniciais no erro de validação), se for satisfeito o critério de *progresso de treinamento* do *Proben1* [Prechelt, 1994], com $P_5(t) < 0.1$, ou se a quantidade máxima de 5.000 iterações for alcançada.

O critério GL_5 fornece uma idéia sobre a perda da generalização durante o treinamento da rede, sendo muito útil para evitar *overfitting* [Prechelt, 1994]. É definido como o aumento do erro de validação em relação ao erro de validação mínimo obtido. Nas Equações 6.1, 6.2 e 6.3, E representa a função de erro quadrático, $E_{tr}(t)$ o erro médio por padrão sobre o conjunto de treinamento, medido durante a época t , $E_{va}(t)$ o erro no conjunto de validação após a época t , e $E_{opt}(t)$ corresponde ao menor erro de validação obtido até a época t , descrito pela Equação (6.1).

$$E_{opt}(t) = \min_{t' \leq t} E_{va}(t') \quad (6.1)$$

A perda de generalização (GL) na época t é definida como o aumento do erro de validação em relação ao menor erro de validação obtido até a época t (em porcentagem). A perda de generalização é descrita pela Equação (6.2).

$$GL(t) = 100.(\frac{E_{va}(t)}{E_{opt}(t)} - 1) \quad (6.2)$$

Desta forma, é estabelecido um limiar α , de modo que o critério GL_α é caracterizado pela parada do treinamento quando a perda de generalização ultrapassa o valor de α . Sendo assim, o critério GL_5 estabelece a parada do treinamento quando $GL(t) > 5$.

O critério de progresso de treinamento do *Proben1* [Prechelt, 1994] fornece uma idéia de quanto o erro médio de treinamento é maior que o mínimo erro de treinamento, durante uma seqüência de épocas consecutivas. Considerando uma seqüência de k épocas numeradas de $n + 1$ até $n + k$, sendo n divisível por k , o progresso de treinamento P (medido em partes por mil) após esta seqüência de k épocas é dado pela Equação (6.3).

$$P_k(t) = 1000.(\frac{\sum_{t' \in t-k+1, \dots, t} E_{tr}(t')}{k. \min_{t' \in t-k+1, \dots, t} E_{tr}(t')}) \quad (6.3)$$

É estabelecido um limiar, de modo que o critério do progresso de treinamento é caracterizado pela parada do treinamento quando o progresso de treinamento fica menor que este limiar.

6.1.5 Resultados obtidos

Para a rede Perceptron Multi-Camadas foram realizados dois experimentos distintos: um experimento onde o método de pré-processamento utilizado é somente a Normalização dos padrões captados pelo sistema sensor do nariz artificial, de modo que estes valores fiquem compreendidos no intervalo $[-1, 1]$; e um experimento que utiliza

um filtro Wavelet como método de pré-processamento, para redução do ruído e compressão dos dados captados pelos sensores do nariz artificial.

Para cada um dos experimentos apresenta-se uma tabela contendo a topologia considerada, os aspectos observados nas trinta (30) execuções realizadas, bem como os valores de média e desvio-padrão encontrados. Em seguida, são tecidos comentários sobre o desempenho da rede neural artificial na resolução do problema abordado. Tais resultados estão publicados em [Zanchettin et al., 2003 e Zanchettin e Ludermir, 2003a].

Na Tabela 6.1 são apresentados os resultados obtidos no treinamento das redes Perceptron Multi-Camadas, com a Partição 1, utilizando somente a Normalização como método de pré-processamento.

Tabela 6.1 – MLP - Média e desvio-padrão para Partição 1

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	2103.74	3918.05	1062.53	1979.60	1085.75	1985.66	0.1434	0.2151	77	88
8	1749.77	2484.75	881.41	1235.83	898.56	1247.39	0.1388	0.1875	59	28
12	3416.32	4939.83	1713.30	2453.92	1739.90	2474.95	0.2186	0.2696	84	57
16	4248.31	4334.83	2117.74	2144.50	2161.37	2170.85	0.3167	0.2965	89	64
20	5694.61	5050.60	2852.58	2527.16	2915.66	2554.76	0.3444	0.2568	94	49
30	6888.90	4532.85	3428.18	2244.65	3489.28	2279.32	0.4365	0.2579	69	37

Para este experimento, o menor erro de classificação obtido dentre todas as topologias testadas foi de 0.22% de erro, obtido pela topologia com 8 nodos na camada intermediária. Esta topologia também apresentou o menor erro médio de classificação, um erro de 13.88%. Como pode ser observado na Tabela 6.1, a topologia com 8 unidades de processamento na camada intermediária apresentou melhor desempenho que as demais topologias em todos os critérios observados nesta análise. Além de uma média menor no número de épocas de treinamento, a topologia com 8 nodos na camada intermediária apresentou o menor desvio-padrão para todos os critérios. O baixo desvio-padrão indica uma maior uniformidade nos resultados.

Para a Partição 2, o menor erro médio de classificação foi obtido pela topologia com 4 nodos na camada intermediária, um erro de 14, 63% para o conjunto de teste. Como pode ser observado na Tabela 6.2, os demais aspectos observados nesta avaliação

não apresentaram diferenças significativas em relação aos resultados encontrados nos experimentos com a Partição 1.

Tabela 6.2 – MLP - Média e desvio-padrão para Partição 2

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	2085.96	3292.44	1070.11	1685.21	1108.01	1656.09	0.1463	0.1742	71	41
8	1885.50	2521.46	980.22	1303.07	1016.87	1292.51	0.1632	0.2139	79	35
12	2860.74	3866.08	1434.28	1943.66	1481.69	1976.48	0.2080	0.2578	139	82
16	3639.09	4511.08	1827.89	2243.12	1873.71	2267.84	0.2460	0.2561	168	131
20	5304.18	5454.14	2658.36	2714.84	2698.73	2703.68	0.3512	0.37091	156	108
30	6383.97	4967.68	3207.20	2469.36	3255.38	2484.26	0.4137	0.2779	204	181

Na Tabela 6.3 são apresentados os resultados dos experimentos realizados com a Partição 3, o menor erro médio de classificação foi obtido pela topologia com 8 nodos da camada intermediária, um erro de 13,65% sobre o conjunto de teste.

Tabela 6.3 – MLP - Média e desvio-padrão para partição 3

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	2085.96	3292.44	1070.11	1685.21	1078.96	1677.71	0.1419	0.1753	71	41
8	1914.03	3039.22	1039.65	1599.26	989.45	1574.15	0.1365	0.1726	73	36
12	2904.99	4515.43	1499.63	2254.86	1442.17	2190.87	0.1805	0.2443	137	94
16	3184.59	3677.86	1661.25	1893.30	1585.36	1831.77	0.2239	0.2351	204	153
20	5909.82	5041.43	3025.91	2522.94	2977.98	2507.81	0.3608	0.2987	147	133
30	7709.14	5412.66	3939.72	2740.37	3870.54	2735.36	0.4543	0.2783	139	151

As Tabelas 6.1, 6.2 e 6.3 também demonstram que a partir da topologia com 8 nodos na camada intermediária, o erro de classificação aumenta com a adição de novas unidades de processamento nesta camada. Apesar deste indício, foram realizados experimentos com topologias onde o número de unidades de processamento na camada intermediária ultrapassava 30 nodos (com 50, 70, 100, 150 e 300 unidades de processamento na camada intermediária). Porém, nestes experimentos o erro de classificação manteve a mesma tendência de alta, proporcional ao número de unidades de processamento adicionadas.

Nas Tabelas 6.4, 6.5 e 6.6 são apresentados os resultados do segundo treinamento realizado com as redes Perceptron Multi-Camadas. Neste experimento a Análise Wavelet é utilizada como um filtro para o ruído presente no sinal captado pelo sistema sensor do nariz artificial e como um mecanismo de compressão para os dados de odor analisados.

Tabela 6.4 – MLP com Wavelet - Média e desvio-padrão para a Partição 1

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	1907.08	2889.74	1006.46	1440.85	966.94	1439.19	0.1366	0.1880	75	35
8	1249.58	2110.51	652.59	1064.61	626.92	1025.68	0.1029	0.1695	108	79
12	2670.21	4056.98	1352.79	2038.80	1337.19	2011.05	0.1826	0.2565	343	429
16	3910.73	5162.13	1979.67	2584.60	1962.24	2572.54	0.2409	0.2811	354	402
20	5250.95	4748.93	2660.83	2376.97	2633.50	2374.60	0.3392	0.2719	452	916
30	6534.00	5664.36	3295.29	2857.57	3276.19	2829.78	0.4060	0.2901	231	260

Tabela 6.5 – MLP com Wavelet - Média e desvio-padrão para a Partição 2

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	1950.06	2870.58	1018.69	1461.95	989.09	1465.36	0.1369	0.2081	83	35
8	1541.33	2468.88	810.58	1262.98	810.36	1272.35	0.1287	0.1938	74	25
12	2668.36	4135.93	1365.18	2090.97	1334.44	2071.76	0.1774	0.2530	205	208
16	3593.02	4640.31	1842.22	2356.56	1829.93	2378.96	0.2383	0.2902	237	222
20	5390.32	4497.10	2736.58	2255.51	2736.40	2283.39	0.3403	0.2628	299	337
30	6313.22	5753.94	3206.66	2894.55	3193.49	2909.74	0.3769	0.3158	279	281

Tabela 6.6 – MLP com Wavelet - Média e desvio-padrão para a Partição 3

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	2344.52	4444.17	1166.67	2230.25	1204.51	2228.21	0.1400	0.2156	83	34
8	1486.09	2542.49	748.16	1299.52	756.73	1284.23	0.1135	0.1571	183	364
12	2582.14	4643.59	1269.52	2318.08	1309.24	2342.77	0.1587	0.2545	262	254
16	3780.49	5029.65	1876.63	2509.66	1908.30	2562.84	0.2302	0.2870	369	480
20	5726.77	5493.84	2856.33	2731.17	2914.56	2807.27	0.3542	0.3154	304	409
30	6496.66	6118.10	3250.62	3037.42	3305.73	3089.69	0.3612	0.2920	342	450

A partir dos resultados apresentados nas Tabelas 6.4, 6.5 e 6.6, se percebe que a utilização da Transformada de Wavelet, como método de pré-processamento para os padrões de odores, melhora consideravelmente o desempenho do método de classificação. O menor erro de classificação observado dentre todas as topologias, um erro de classificação zero, foi obtido pelas topologias com 20 e 30 unidades de processamento na camada intermediária, nas três partições testadas. Apesar do bom desempenho, na média de classificação das 30 execuções, estas topologias ficaram muito acima dos valores obtidos pela topologia com 8 nodos na camada intermediária. O menor erro médio de classificação foi obtido pela topologia com 8 nodos na camada intermediária (Tabela 6.4), um erro de 10.29% para o conjunto de teste. O menor erro de classificação obtido pela topologia com 8 nodos na camada intermediária também foi baixo, um erro de 0.07%, na Partição 1.

Da mesma forma que no experimento anterior, o erro de classificação tende a aumentar com o acréscimo de unidades de processamento na camada intermediária. Os mesmos experimentos relatados na descrição do experimento anterior (inserção de um maior número de unidades de processamento na camada intermediária) foram realizados e seus resultados demonstram comportamento similar aos resultados obtidos pela rede Perceptron Multi-Camadas sem Pré-processamento Wavelet.

Se compararmos os melhores resultados obtidos pelos dois experimentos (resultados obtidos pelas topologias com 8 nodos na camada intermediária, com e sem a utilização do pré-processamento com a Transformada de Wavelet) podemos verificar que a melhora foi substancial. A medida do erro de classificação apresentado no primeiro experimento (Tabela 6.1, 6.2 e 6.3) foi de 14.61% e no segundo de 11.50% (Tabela 6.4, 6.5 e 6.6).

A fim de verificarmos a relevância estatística dos resultados obtidos, é apresentado na Tabela 6.7 o teste de hipóteses para estes experimentos. O teste é realizado a partir dos melhores resultados médios, obtidos por cada um dos experimentos (sem e com pré-processamento com Análise Wavelet). Neste teste, verificamos se o desempenho do classificador nos dois experimentos é similar ou não.

Tabela 6.7 - Estatística do teste emparelhado *t* de Student

$H_0 : \mu_1 = \mu_2$	
$H_1 : \mu_1 \neq \mu_2$	
Valor calculado da variável de teste	Decisão
-0.6591	Não rejeitada

A Tabela 6.7 apresenta os resultados do teste *t* para amostras emparelhadas, com nível de significância de 5%. Nesta tabela, μ_1 e μ_2 são respectivamente as médias dos experimentos com redes Perceptron Multi-Camadas sem o uso do pré-processamento com a Análise Wavelet e as médias dos experimentos utilizando o pré-processamento com a Análise Wavelet. Apesar da menor média de classificação do segundo experimento, estatisticamente os resultados comprovam a hipótese de que a performance média das redes Perceptron Multi-Camadas fazendo uso da Análise Wavelet é tão boa quanto à das redes Perceptron Multi-Camadas que não usam este tipo de pré-processamento.

6.2 Experimentos com a Rede Neural com Atrasos no Tempo

A abordagem utilizada nestes experimentos foi proposta por Yamazaki (2001). Esta abordagem consiste de um sistema de reconhecimento de padrões capaz de analisar as características temporais dos sinais gerados pelos sensores do nariz artificial, levando em consideração a variação destes valores ao longo do intervalo de tempo em que foram feitas as aquisições dos dados.

6.2.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste

Como descrito em Yamazaki (2001), a arquitetura do sistema de reconhecimento de padrões necessita receber curvas completas geradas pelos sensores do nariz artificial durante a aquisição dos dados. Tais curvas precisam ser obtidas em intervalos

ininterruptos de tempo, desta forma, apenas os valores obtidos em cada uma das repetições podem ser utilizados no conjunto de treinamento, validação e teste (os padrões não podem ser retirados da ordem de aquisição ou agrupados em conjuntos maiores).

Desta forma, uma das repetições da aquisição dos dados (960 padrões de cada um dos cinco gases) foi utilizada como conjunto de treinamento. Da mesma forma, outras duas repetições distintas (com a mesma quantidade de dados) foram utilizadas como conjunto de validação e teste. A escolha das repetições foi feita de forma arbitrária utilizando-se as três primeiras repetições da aquisição dos dados, para que a ordem de aquisição e apresentação dos dados não interferisse nos resultados do experimento.

Para a formação das 3 partições de dados utilizadas nos experimentos, as demais aquisições de dados foram utilizadas. Assim, como existem nove repetições de aquisições, todas foram utilizadas no treinamento da rede, seja como conjunto de treinamento, validação ou teste.

Foi utilizada a mesma base de dados que gerou os resultados da seção anterior, para as redes Perceptron Multi-Camadas, permitindo assim uma melhor comparação entre as abordagens utilizadas.

6.2.2 Pré-processamento

Utilizou-se a mesma normalização adotada nos experimentos com redes Perceptron Multi-Camadas, desta forma, dois experimentos foram realizados: no primeiro experimento os dados foram pré-processados de modo que seus valores ficassem compreendidos na faixa entre -1 e 1 ; no segundo experimento se utilizou a Transformada de Wavelet como método de pré-processamento.

6.2.3 Topologia utilizada

Para o tratamento temporal dos padrões de odores, foi utilizada uma Rede Neural com Atrasos no Tempo [Lang e Hilton, 1988], contendo apenas uma camada intermediária. A camada de entrada possui 16 unidades de processamento, pelo fato da utilização do atraso no tempo (*delay de $t-1$*), para cada uma das oito curvas de entrada, formadas pelos padrões adquiridos no sistema sensor. Este atraso é formado pelo valor atual de resistência do sensor e por um valor temporalmente anterior a este, de cada um dos sensores. Esta quantidade de atrasos no tempo para cada uma das curvas de entrada foi escolhida por ser a menor configuração possível utilizando características temporais.

A camada de saída possui cinco unidades de processamento, as quais correspondem a cada um dos gases a serem classificados, utilizando a mesma codificação da rede Perceptron Multi-Camadas descrita na seção anterior. Os nodos da camada intermediária e de saída possuem função de ativação *tangente hiperbólica* [Prechelt, 1994]. A rede possui todas as possíveis conexões de propagação direta (*feedforward*) entre as camadas adjacentes, sem possuir conexões que ligam nodos de camadas não-adjacentes. Foram treinadas seis topologias diferentes (com 4, 8, 12, 16, 20 e 30 unidades de processamento na camada intermediária).

6.2.4 Metodologia de treinamento

O algoritmo de treinamento utilizado nestes experimentos é similar ao adotado para os experimentos com as redes Perceptron Multi-Camadas na seção anterior, permitindo uma comparação mais justa entre as abordagens. Da mesma forma, foram realizadas trinta (30) execuções com inicializações dos valores das conexões entre as camadas de processamento, diferentes e aleatórias. A quantidade máxima de iterações escolhida foi de 5.000 épocas de treinamento. Os critérios de parada do treinamento foram os mesmos adotados nos experimentos com Perceptron Multi-Camadas.

É importante ressaltar que os erros analisados foram computados separadamente para cada gás, ou seja, a rede avaliou, para cada sensor, as cinco curvas correspondentes aos cinco gases independentemente, e não uma única curva contendo os valores dos

cinco gases em seqüência. Isto foi adotado para que a ordem de apresentação dos gases à rede não influencie nos resultados obtidos.

6.2.5 Resultados obtidos

Como descrito anteriormente, foram realizados dois experimentos distintos: em um experimento foi utilizada a Normalização, para que os valores reais dos dados ficassem compreendidos entre o intervalo $[-1, 1]$; e um experimento que utiliza um filtro Wavelet como método de pré-processamento, para redução do ruído e compressão dos dados capturados pelos sensores do nariz artificial.

Para a análise dos resultados é apresentada uma tabela contendo a topologia considerada, os aspectos observados nas 30 execuções realizadas, bem como os valores de média e desvio-padrão encontrados. Na seqüência são tecidos comentários sobre o desempenho da rede neural artificial, na resolução do problema de classificação de odores investigado. Estes resultados estão publicados em [Zanchettin et al., 2003 e Zanchettin e Ludermir, 2003a].

Nas Tabelas 6.8, 6.9 e 6.10 são apresentados os resultados do treinamento obtido com as Redes Neurais com Atrasos no Tempo, para as três partições testadas. Nestes experimentos somente foi utilizada a Normalização, como método de pré-processamento.

Tabela 6.8 – TDNN - Média e desvio-padrão para a Partição 1

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	7.99	5.45	85.18	24.04	62.35	19.05	0.1075	0.0919	12	9
8	3.22	2.88	75.61	28.83	63.99	15.94	0.0969	0.0881	18	21
12	5.13	10.35	76.11	33.04	67.46	14.44	0.2032	0.1048	17	21
16	48.11	53.58	107.85	49.02	89.13	39.54	0.3166	0.1637	10	3
20	56.55	55.49	115.61	46.28	98.09	53.65	0.3391	0.1751	13	16
30	56.81	49.17	114.10	41.28	99.08	44.26	0.3066	0.1821	10	2

Tabela 6.9 – TDNN - Média e desvio-padrão para a Partição 2

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	5,73	3,90	84,49	26,71	61,75	19,30	0,1362	0,1075	16	23
8	2,72	2,95	71,63	31,13	63,92	16,59	0,1261	0,0872	32	37
12	4,52	9,75	70,41	34,69	70,33	15,82	0,2061	0,1055	33	36
16	47,57	50,21	103,19	43,19	90,44	42,32	0,3266	0,1695	17	23
20	45,65	0,14	98,32	37,04	100,27	44,63	0,3060	0,1834	18	23
30	46,94	34,33	96,46	30,05	98,18	39,09	0,3012	0,1720	11	3

Tabela 6.10 – TDNN - Média e desvio-padrão para a Partição 3

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	5,75	3,89	85,42	26,31	61,85	19,51	0,1341	0,1077	16	23
8	2,81	2,92	72,90	30,60	64,14	16,38	0,1173	0,0865	27	32
12	4,66	9,86	71,83	34,23	69,91	15,57	0,2062	0,1056	29	35
16	44,72	45,58	107,41	50,45	84,00	33,16	0,3021	0,1304	17	23
20	54,11	50,63	106,35	47,76	104,24	52,98	0,3262	0,2032	17	23
30	46,62	35,76	98,07	27,82	98,67	31,79	0,3302	0,1624	11	3

Nestes experimentos, o menor erro de classificação obtido entre as topologias testadas foi 0.16%, obtido na maioria das topologias. Como pode ser observado, a abordagem temporal diminuiu consideravelmente a variação dos resultados, refletido no baixo desvio-padrão apresentado pelos experimentos (Tabelas 6.8, 6.9 e 6.10).

A topologia com menor erro de classificação neste experimento foi a topologia com 8 unidades de processamento na camada intermediária. Nos experimentos realizados com a Partição 1 (Tabela 6.8), o erro de classificação da topologia com 8 nodos foi de 9.69%. Para as Partições 2 e 3 o menor erro médio de classificação também foi obtido pela topologia com 8 nodos na camada intermediária, um erro médio de 12.61% (Tabela 6.9) e 11.73% (Tabela 6.10), respectivamente. É importante salientar, ainda, que os conjuntos de treinamento, validação e teste são formados por diferentes repetições de aquisição de dados, e as características das respostas dos sensores se modificam de uma repetição para a outra [Santos, 2000]. A partir desta afirmação, pode-

se perceber que a rede conseguiu generalizar muito bem o conhecimento adquirido em seu treinamento.

Outro fato a ser observado nesta análise, da mesma forma que o ocorrido com as redes Perceptron Multi-Camadas, o acréscimo no número de unidades de processamento na camada intermediária não melhora o desempenho da rede. Nos testes realizados, com a adição de unidades de processamento na camada intermediária o erro de classificação apresentado pela rede mantém-se estável, na faixa de 30% de erro.

Nas Tabelas 6.11, 6.12 e 6.13 são apresentados os resultados do segundo treinamento realizado com as Redes Neurais com Atrasos no Tempo, utilizando a Análise Wavelet como um filtro para o ruído e como mecanismo de compressão para os sinais de odores.

Tabela 6.11 – TDNN com Wavelet - Média e desvio-padrão para a Partição 1

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	1.27	1.85	31.34	24.45	70.51	10.96	0.0258	0.0615	253	238
8	1.34	1.98	27.17	21.87	66.49	6.99	0.0090	0.0383	192	229
12	3.87	9.63	36.07	23.67	70.16	10.93	0.0200	0.0610	198	233
16	39.61	41.85	74.79	49.89	86.85	35.76	0.1889	0.1706	94	184
20	62.11	55.56	96.19	65.45	98.36	41.77	0.2917	0.1867	60	149
30	61.55	39.16	87.62	43.59	98.34	44.30	0.3003	0.2151	28	89

Tabela 6.12 – TDNN com Wavelet - Média e desvio-padrão para a Partição 2

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	1.13	1.69	30.36	24.18	71.87	13.28	0.0348	0.0722	264	242
8	1.19	1.96	26.42	22.10	66.82	7.54	0.0068	0.0365	225	240
12	3.78	9.64	33.43	23.74	70.76	11.26	0.0200	0.0610	229	241
16	51.13	56.93	83.87	59.24	95.24	49.63	0.2135	0.2149	45	124
20	42.30	38.90	74.08	47.55	89.15	36.99	0.2133	0.2097	111	198
30	53.93	39.17	84.25	41.40	89.32	44.24	0.2754	0.1747	28	89

Tabela 6.13 – TDNN com Wavelet - Média e desvio-padrão para a Partição 3

Número de nodos	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
4	1.14	1.68	30.37	24.23	71.89	13.17	0.0341	0.0698	264	243
8	1.23	1.97	26.36	22.05	66.93	7.63	0.0067	0.0365	207	236
12	3.78	9.64	33.67	23.71	70.83	11.32	0.0200	0.0610	229	241
16	34.17	48.19	71.54	48.78	79.89	29.47	0.1661	0.1820	79	168
20	45.29	38.56	76.50	42.64	88.50	35.56	0.2065	0.1856	110	198
30	40.86	36.74	68.26	40.37	86.88	33.45	0.2067	0.1780	94	185

Como ocorreu nos experimentos com a rede Perceptron Multi-Camadas, a utilização da Transformada de Wavelet como método de pré-processamento melhorou os resultados apresentados pela rede neural artificial. No caso das Redes Neurais com Atrasos no Tempo esta melhora foi considerável. O melhor resultado dos experimentos sem pré-processamento Wavelet foi um erro médio de classificação de 9.69% (Tabela 6.8). Nos experimentos com pré-processamento Wavelet o mais baixo erro médio de classificação foi de 0.67%, obtido pela topologia com 8 unidades de processamento na camada intermediária, nos experimentos com a Partição 3 (Tabela 6.13).

Nos experimentos descritos pelas Tabelas 6.11, 6.12 e 6.13, todas as topologias apresentaram em seus treinamentos erros de classificação nulos, porém apenas as topologias com 4, 8 e 12 unidades de processamento na camada intermediária apresentaram melhora média significativa em relação aos experimentos sem pré-processamento Wavelet.

Apesar da melhora perceptível apresentada pelo segundo experimento (com pré-processamento Wavelet) realizamos o teste de hipótese para verificar se esta melhora estatisticamente se confirma. Na Tabela 6.14 é apresentado o teste de hipóteses para os experimentos acima descritos. Neste teste verificamos se o desempenho da Rede Neural com Atrasos no Tempo com pré-processamento Wavelet é superior ao desempenho da mesma rede sem o pré-processamento.

Tabela 6.14 - Estatística do teste emparelhado de *Student*

$H_0 : \mu_1 = \mu_2$	
$H_1 : \mu_1 > \mu_2$	
Valor calculado da variável de teste	Decisão
-6.2713	Rejeitada

A Tabela 6.14 apresenta os resultados do teste t para amostras emparelhadas, com nível de significância de 5%. Nesta tabela, μ_1 e μ_2 são respectivamente as médias dos experimentos com as Redes Neurais com Atrasos no Tempo sem o uso do pré-processamento com Análise Wavelet e as médias dos experimentos utilizando o pré-processamento com Análise Wavelet. A partir do teste de hipótese se pode concluir que a performance média das Redes Neurais com Atrasos no Tempo fazendo uso da análise Wavelet é melhor que a das Redes Neurais com Atrasos no Tempo que não utilizam este tipo de pré-processamento.

A partir dos resultados obtidos pela abordagem temporal, pode-se concluir que a utilização do método de pré-processamento com filtro Wavelet faz das redes neurais artificiais uma ferramenta poderosa na classificação de padrões de odores.

6.3 Experimentos com a rede *Feature-weighted Detector*

Este modelo oferece especial interesse no tratamento dos padrões de odores por se propor a resolver dois dos principais problemas no reconhecimento de padrões: a seleção de atributos e a classificação dos padrões, além da capacidade de extrair regras difusas do tipo *Se-Então* da rede. Nas próximas seções serão descritos os experimentos e resultados encontrados no treinamento da rede *Feature-weighted Detector* [Li et al., 2002].

6.3.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste

A partir de experimentos prévios, realizados com a rede *Feature-weighted Detector*, verificou-se que a ordem com que os padrões de odores são apresentados no treinamento, possui influência sobre seu desempenho na classificação dos gases. Assim,

nestes experimentos, os exemplos de cada um dos gases foram apresentados de forma ordenada no treinamento da rede (primeiramente foram apresentados todos os exemplos de um tipo de gás, posteriormente de outro tipo de gás, etc.). Para o teste da rede neurodifusa os padrões de odores foram apresentados de forma aleatória.

A fim de considerar esta necessidade de ordenação e verificar o desempenho da rede Feature-weighted Detector com várias partições de dados, os mesmos conjuntos empregados nos experimentos com a Rede Neural com Atrasos no Tempo foram utilizados. Desta forma, cada uma das partições é composta por duas repetições distintas das aquisições dos dados: uma das repetições da aquisição dos dados (960 padrões de cada um dos cinco gases) foi utilizada como conjunto de treinamento, com exemplos ordenados por classe; e outra repetição (com a mesma quantidade de dados) foi utilizada como conjunto de teste, com exemplos dispostos de forma aleatória.

A rede Feature-weighted Detector não utiliza conjunto de validação em seu treinamento. É importante salientar, que os conjuntos de treinamento e teste de cada uma das partições são formados por diferentes repetições de aquisições de dados, e as características das respostas dos sensores se modificam de uma repetição para outra [Santos, 2000].

6.3.2 Pré-processamento

O conjunto de padrões foi normalizado de modo a ficar compreendido na faixa de valores entre 0 e 1. O pré-processamento a partir da Transformada de Wavelets não foi utilizado nos experimentos com a rede Feature-weighted Detector, pois em análises preliminares não melhoraram os resultados apresentados pelo modelo. Por questão de objetividade foram suprimidos dos resultados obtidos com a Transformada de Wavelet no treinamento da rede Feature-weighted Detector.

6.3.3 Topologia utilizada

Nos experimentos realizados com a rede Feature-weighted Detector, foram utilizadas oito unidades de processamento de entrada, que correspondem aos atributos dos padrões de odores. E cinco unidades de saída, relacionadas a cada um dos gases a serem classificados. Neste experimento foi utilizada a codificação *1-of-m*, onde m é o número de classes [Prechelt, 1994].

6.3.4 Metodologia de treinamento

Foram realizadas 30 (trinta) execuções com a rede Feature-weighted Detector, para cada uma das partições. Foram utilizadas inicializações baseadas nos centros dos dados para as conexões de memória e valores fixos em 1 para as conexões de pesos. Os critérios de parada do treinamento utilizados foram o erro mínimo de treinamento ε e o número máximo de 5.000 iterações.

6.3.5 Resultados obtidos

Para cada uma das partições, os aspectos observados nas 30 execuções realizadas, bem como os valores de média e desvio-padrão são apresentados. Em seguida, são tecidos comentários sobre o desempenho da rede na resolução do problema abordado. Estes resultados estão publicados em [Zanchettin e Ludermir, 2003b].

Na Tabela 6.15 são apresentados os resultados obtidos no treinamento da rede Feature-weighted Detector, na classificação dos cinco gases, em cada uma das partições.

Tabela 6.15 - Média e desvio-padrão dos resultados obtidos com a rede Featureweighted Detector

Partição	Erro de Treinamento		Erro de Teste		Erro de Classificação		Épocas Treinadas	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
1	134.35	5.78	570.99	36.92	0.2218	0.06	5000	0
2	177.60	0.01	430.00	0.00	0.2000	0.00	5000	0
3	177.62	0.05	430.00	0.00	0.2000	0.00	5000	0

Para esta abordagem o melhor erro médio de classificação obtido foi de 20%, obtido nas partições 2 e 3. Um erro relativamente alto se levarmos em consideração as abordagens previamente apresentadas. O menor erro de classificação obtido nos experimentos com a rede Feature-weighted Detector foi de 20%, encontrado na grande maioria das execuções. Verificou-se que a rede apresenta dificuldades em separar os gases Etano e Metano, indicados na Análise de Componentes Principais como classes muito próximas. Isto pode caracterizar uma deficiência da rede Feature-weighted Detector em tratar dados não-linearmente separáveis.

Pode-se verificar também Tabela 6.15, que o critério de parada em todos os experimentos foi o número máximo de iterações. A partir desta constatação tentou-se modificar o treinamento da rede, de modo a inserir um critério de parada, baseado na utilização de um conjunto de validação. Porém, os resultados apresentados não foram satisfatórios, pois, oscilações iniciais no treinamento da rede faziam com que o treinamento fosse interrompido de maneira prematura.

Outra alteração realizada sobre a rede Feature-weighted Detector original foi quanto a inicialização das conexões de pesos $\{w_{ij}\}$. As conexões de pesos originalmente são inicializadas com valores iguais a 1, proposta por Li et al. (2001). Foram realizados experimentos onde estas conexões eram inicializadas com valores aleatórios [Amorin et al., 2003]. A partir dos resultados, verificou-se que as duas abordagens obtêm desempenhos similares.

Nas Tabelas 6.16 e 6.17, são apresentadas as médias dos valores das conexões de memória e peso, obtidas após o treinamento da rede com a Partição 1. Estes valores são utilizados na seleção dos atributos relevantes e na extração de regras difusas da rede.

Tabela 6.16 - Médias dos valores finais para as conexões de memória

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
m_1	0.982573	0.961906	0.960222	0.971907	0.949086	0.946496	0.959430	0.680431
m_2	0.690826	0.527188	0.463760	0.036989	0.486119	0.542842	0.298441	0.372369
m_3	0.593980	0.497958	0.390256	0.072632	0.408675	0.506118	0.535367	0.235794
m_4	0.702855	0.528086	0.465679	0.038395	0.487341	0.546711	0.304195	0.370047
m_5	0.755790	0.552057	0.530144	0.004374	0.553913	0.570328	0.084244	0.497325

Tabela 6.17 - Médias dos valores finais para as conexões de pesos

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
w_1	0.913683	0.837864	0.805750	0.418144	0.838599	0.876073	0.702558	0.911053
w_2	0.994720	0.900508	0.998670	0.999827	0.998891	0.900000	0.099999	0.998007
w_3	0.000000	0.000000	0.001753	0.999995	0.000004	0.000000	0.999817	0.999978
w_4	0.935550	0.977357	0.765897	0.000000	0.730483	0.992370	0.089121	0.076219
w_5	0.000000	0.001045	0.000000	0.998079	0.000000	0.000000	0.924484	0.949968

6.3.5.1 Seleção de Atributos

No modelo Feature-weighted Detector a seleção de atributos é realizada a partir dos valores das conexões de pesos obtidas durante a fase de treinamento da rede. Tais valores representam o grau de contribuição dos atributos para cada uma das classes.

Como a base de dados utilizada nos experimentos foi dividida em três partições distintas, os valores resultantes das conexões de pesos para cada uma das partições variam. Por esta razão, a utilização da média dos valores das conexões nas três partições, como meio de escolha da relevância dos atributos, prejudicaria a análise do método de seleção. Desta forma, uma das partições foi escolhida de forma arbitrária, para aplicação e avaliação do método de seleção de atributos da rede. A partição escolhida, de maneira arbitrária, para os experimentos foi a Partição 1.

A Tabela 6.18 apresenta os resultados de dois experimentos, realizados com a Partição 1. No primeiro, a rede foi treinada com os oito atributos da base de dados. Considerando-se estes resultados, pode-se observar que: (1) os atributos x_2 , x_3 , e x_5 possuem pequena contribuição na formação dos agrupamentos das classes 3 e 5 e o atributo x_4 possui pequena participação nos agrupamentos das classes 1 e 4; e (2) a contribuição dos atributos x_2 , x_3 , x_4 e x_5 para as classe 1, 3 e 4 pode ser desempenhada pelos demais atributos, que apresentaram grande contribuição na formação dos agrupamentos destas classes.

Tabela 6.18 - Resultados da rede *Feature-weighted Detector* na classificação de sinais de odores

Parametros	Experimento 1 $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$								Experimento 2 $\{x_1, x_2, x_3, x_4\}$			
σ	0.225								0.225			
α	0.004								0.004			
β	0.3								0.3			
Média	0.2218								0.2016			
Desvio-padrão	0.06								0.00			
w_1	0.913	0.837	0.805	0.418	0.835	0.876	0.702	0.911	1.000	0.613	0.810	0.999
w_2	0.994	0.900	0.998	0.999	0.998	0.900	0.099	0.998	0.999	0.999	0.417	0.999
w_3	0.000	0.000	0.001	0.999	0.000	0.000	0.999	0.999	0.000	0.000	0.999	0.999
w_4	0.935	0.977	0.765	0.000	0.730	0.992	0.089	0.076	0.999	0.894	0.004	0.000
w_5	0.000	0.001	0.000	0.998	0.000	0.000	0.924	0.949	0.000	0.000	0.899	0.940

Na Tabela 6.18, σ representa a nebulosidade no agrupamento, α representa a taxa de aprendizado temporal e β a taxa de aprendizado. A letra x representa os atributos dos padrões de odores e w a média das conexões de pesos.

No segundo experimento, os atributos x_2 , x_3 , x_4 e x_5 foram excluídos da base de dados e a rede foi novamente treinada somente com os quatro atributos restantes (x_1 , x_6 , x_7 e x_8). No treinamento, foi utilizada a mesma configuração de parâmetros do primeiro experimento. Mesmo com a retirada dos referenciados atributos, a média do erro de classificação da rede se manteve estável. No segundo experimento, a média do erro de classificação obtido foi de 20.16%, inferior ao primeiro experimento onde o erro médio era de 22.18%. Este resultado comprova a pequena relevância dos atributos x_2 , x_3 , x_4 e x_5 na formação dos agrupamentos. Os sensores indicados como de menor relevância foram: Octil-sulfonato de Sódio, Ácido Nitrobenzeno Sulfônico, Antraquinona Sulfato de Sódio e Ácido P-toluenosulfônico.

A fim de comprar a eficiência da seleção de atributos da rede *Feature-weighted Detector*, foram realizados experimentos com os demais modelos apresentados nesta dissertação. Na Tabela 6.19 são apresentados os experimentos com todos os classificadores utilizados na dissertação: Perceptron Multi-Camadas (MLP), Redes Neurais com Atrasos no Tempo (TDNN), *Feature-weighted Detecto* (FWD) e *Evolving Fuzzy Neural Networks* (EFuNN). São realizados dois experimentos, um com todos os atributos da base de dados e outro excluindo-se os atributos indicados pela rede *Feature-*

weighed Detector como de menor relevância para a formação dos agrupamentos de dados. Os dados utilizados para este experimento são provenientes da Partição 1.

Tabela 6.19 - Comparação dos resultados da seleção de atributos

Modelo	Erro de classificação para o conjunto de teste, base com 8 atributos		Erro de classificação para o conjunto de teste, base com 4 atributos	
	Média	Desvio	Média	Desvio
MLP	0.1388	0.1874	0.1402	0.1952
TDNN	0.0969	0.0881	0.0982	0.0754
FWD	0.2218	0.0600	0.2016	0.0000
EFuNN	0.0088	0.0028	0.0333	0.0035

Novamente pôde ser verificado que, mesmo com a exclusão dos quatro atributos da base de dados (x_2 , x_3 , x_4 e x_5) o erro de classificação apresentado pelos quatro modelos testados manteve-se estável. Outra questão a ser ressaltada é que a rede Feature-weighted Detector indicou como de menor relevância atributos também indicados pela Análise de Componentes Principais, como de menor relevância, na análise descrita pela Tabela 3.4.

6.3.5.2 Extração de regras

As regras extraídas do modelo Feature-weighted Detector são regras explicativas difusas do tipo *Se...Então*, onde a parte *Se* da regra se refere ao conjunto de atributos com valores previsores, e a conclusão da regra *Então* especifica o valor esperado da classe.

A extração das regras difusas *Se-Então* da rede Feature-weighted Detector é realizada a partir das conexões de memória obtidas durante a fase de treinamento. Os valores destas conexões de memória são apresentados na Tabela 6.16.

Como descrito em Li et al. (2001), as seguintes regras difusas podem ser construídas a partir das conexões de memória, obtidas pelo treinamento da rede Feature-weighted Detector com os padrões de odores:

Regra₁ Se x_1 é aproximadamente 0.9802 E x_2 é aproximadamente 0.9432 E x_3 é aproximadamente 0.9555 E x_4 é aproximadamente 0.6767
Então é Butano.

Regra₂ Se x_1 é aproximadamente 0.6712 E x_2 é aproximadamente 0.5356 E x_3 é aproximadamente 0.3334 E x_4 é aproximadamente 0.3519
Então é Etano.

Regra₃ Se x_1 é aproximadamente 0.5973 E x_2 é aproximadamente 0.5149 E x_3 é aproximadamente 0.5427 E x_4 é aproximadamente 0.2479
Então é Monóxido de Carbono.

Regra₄ Se x_1 é aproximadamente 0.7448 E x_2 é aproximadamente 0.5768 E x_3 é aproximadamente 0.3038 E x_4 é aproximadamente 0.4062
Então é Propano.

Regra₅ Se x_1 é aproximadamente 0.7507 E x_2 é aproximadamente 0.5689 E x_3 é aproximadamente 0.0882 E x_4 é aproximadamente 0.4956
Então é Metano.

6.3.5.3 Considerações

A fim de verificar a efetividade do mecanismo de seleção de atributos da rede Feature-weighted Detector, um teste estatístico com os resultados dos experimentos com todos os atributos da base de dados e sem quatro destes atributos, indicados como de menor relevância pelo método é apresentado na Tabela 6.20.

Tabela 6.20 - Estatística do teste emparelhado de *Student*

Modelo	$H_0 : \mu_1 = \mu_2$	
	$H_1 : \mu_1 \neq \mu_2$	
	Valor calculado da variável de teste	Decisão
MLP	0.0749	Não Rejeitada
TDNN	0.7302	Não Rejeitada
FWD	1.4984	Não Rejeitada
EFuNN	-46.7339	Rejeitada

A Tabela 6.20 apresenta os resultados do teste t para amostras emparelhadas, com nível de significância de 5%. Nesta tabela, μ_1 e μ_2 são respectivamente as médias dos experimentos das quatro abordagens investigadas com e sem os atributos x_2 , x_3 , x_4 e

x_5 . Estes resultados comprovam a hipótese de que a performance dos modelos sem os referidos atributos é tão boa quanto o das redes treinadas com todos os atributos da base de dados. Somente a rede Evolving Fuzzy Neural Networks (EFuNN) teve sua média de classificação prejudicada com a seleção de atributos, talvez isto se deva pelo fato da rede utilizar aprendizado local.

Apesar da baixa performance da rede Feature-weighted Detector na classificação dos padrões de odores, ela apresentou promissores resultados na seleção dos atributos mais relevantes para a formação dos agrupamentos. Sua utilização possibilitou uma redução na dimensão da base de padrões de odores, e conseqüentemente, uma redução da complexidade da rede neural artificial utilizada, na classificação dos odores, na ordem de 50%. Isto fica mais evidente se levarmos em consideração que a complexidade da tarefa de reconhecimento de odores é diretamente proporcional ao número de sensores utilizados no reconhecimento do odor [Yamazaki, 2001].

A partir da seleção dos atributos pode-se verificar a relevância de cada um dos sensores no reconhecimento do odor analisado, o que facilita a tarefa de reconhecimento de padrões em problemas específicos. Por exemplo, para detecção de um odor específico, somente os sensores importantes na detecção deste gás em particular são necessários. Os demais sensores podem ser eliminados, diminuindo os custos na construção e a complexidade de operação do sistema.

6.4 Experimentos com a rede Evolving Fuzzy Neural Networks

O interesse na aplicação da rede Evolving Fuzzy Neural Network [Kasabov, 1999] no tratamento dos padrões de odores surgiu a partir de duas de suas funcionalidades: (1) possibilidade de extração do conhecimento da rede a partir de regras de classificação; e (2) treinamento *on-line*, o que possibilita a apresentação de novos padrões e classes a rede, sem a necessidade de re-apresentar a base de dados toda. Em um dispositivo onde

se prima pela eficiência, rapidez, plasticidade e tamanho estas características são muito apreciadas.

6.4.1 Divisão do conjunto de padrões em Treinamento, Validação e Teste

Pelo fato da rede Evolving Fuzzy Neural Network não utilizar conjunto de validação em seu treinamento, a base de dados contendo os 5.400 padrões de odores foi dividida em dois conjuntos distintos de dados: conjunto de treinamento para a rede neuro-difusa, correspondendo a 75% da quantidade total de padrões; e conjunto de teste, correspondendo aos 25% restantes. Da mesma forma que nos demais experimentos, três partições de dados foram utilizadas no treinamento da rede. Cada partição contém a mesma quantidade de padrões de odores, a única diferença entre elas é a ordem de apresentação dos dados.

É importante ressaltar que o conjunto de treinamento possui 75% dos padrões de cada um dos gases (para cada gás, 450 padrões são escolhidos de forma aleatória entre os 600 padrões de odores existentes). O mesmo procedimento é realizado com o conjunto de teste, de forma que os cinco gases sejam representados pela mesma quantidade de padrões em cada um dos conjuntos de dados.

6.4.2 Pré-processamento

A rede Evolving Fuzzy Neural Network não necessita de pré-processamento, como descrito na seção 4.2.2 da presente dissertação, os dados são apresentados à camada de entrada sem pré-processamento algum. Nesta camada os exemplos são transformados em padrões difusos, representados pela segunda camada de processamento da rede, no espaço de representação do problema. Desta forma, os padrões de odores foram apresentados à rede com a mesma amplitude de valores captada pelo sistema sensor do nariz artificial.

6.4.3 Topologia utilizada

Nos experimentos realizados, foram utilizadas oito unidades de processamento de entrada, que correspondem aos atributos dos padrões de odores. Cada unidade de processamento de entrada e saída contém duas funções de pertinência, utilizadas na representação difusa dos dados. Uma unidade de saída foi utilizada, onde cada classe foi representada por números reais, dispostos de 1 a 5, relacionados a cada um dos gases a serem classificados.

6.4.4 Metodologia de treinamento

A inexistência de aleatoriedade no treinamento da rede Evolving Fuzzy Neural Network faz com que uma execução da rede seja suficiente para fornecer uma idéia sobre a generalização do modelo. Ou seja, como o aprendizado é realizado a partir dos exemplos dos dados apresentados ao modelo, os mesmos dados irão convergir sempre para um mesmo ponto no espaço de busca. Assim, para realizar a comparação da rede Evolving Fuzzy Neural Networks com as demais abordagens investigadas neste trabalho foram utilizadas 30 partições, de forma a se obter um erro médio de classificação da rede para a base de padrões de odores utilizada.

O modo de treinamento utilizado foi *one pass training*, os exemplos são apresentados à rede somente uma vez para cada uma das modalidades e partições testadas. Foram realizadas quatro modalidades de experimentos: (1) no primeiro experimento a rede foi inicializada sem nenhuma regra e em seu treinamento não foram realizadas operações de poda (*pruning*) ou agregação de nodos de regras; (2) no segundo experimento, a rede também foi inicializada sem nenhuma regra, mas durante seu treinamento foi utilizada a técnica de poda dos nodos de regras que estivessem fora do limiar de ativação pré-estabelecido; (3) no terceiro experimento, a rede foi inicializada com algumas regras, remanescentes de treinamentos prévios e novos dados foram apresentados à rede de maneira on-line. Durante este treinamento foram utilizados ambos, poda e agregação; e (4) no quarto experimento, a partir de uma arquitetura previamente treinada com quatro gases, foram apresentados exemplos de um

quinto gás, forçando a rede a se adaptar a estes novos exemplos e englobar automaticamente o quinto gás em sua topologia.

Pelo fato do modo de treinamento ser *one pass training*, a parada do treinamento ocorre após a apresentação do último exemplo do conjunto de treinamento.

6.4.5 Resultados obtidos

Nesta seção são apresentados os resultados dos experimentos realizados com a rede Evolving Fuzzy Neural Network, para cada uma das partições testadas. São apresentados os aspectos observados em cada uma das simulações e tecidos comentários sobre o desempenho da rede neuro-difusa na resolução do problema de classificação de odores. Estes resultados estão publicados em [Zanchettin e Ludermir, 2003c].

Na Tabela 6.21 são apresentados os resultados do primeiro experimento, onde uma arquitetura Evolving Fuzzy Neural Network inicializada sem nenhuma regra é treinada com os padrões de odores das três partições testadas. Neste experimento, optou-se por não realizar poda ou agregação sobre os nodos de regras criados na camada de regras durante o treinamento. Desta forma, as três partições foram treinadas a partir da seguinte configuração de parâmetros, obtida após vários treinamentos: *valor do limiar de sensibilidade*, ***sthr*** = 0.9; *valor do limiar de erro*, ***errthr*** = 0.1; *número de funções de pertinência*, ***mf*** = 2; *taxa de aprendizado para $W1$* , ***lr1*** = 0.01; *taxa de aprendizado para $W2$* , ***lr2*** = 0.01; *parâmetro de controle de poda*, ***nodeage*** = 10; *parâmetro de controle de agregação*, ***aggintev*** = -1; *campo máximo de auto-ajuste*, ***maxfield*** = 0.5; *limiar mínimo para extração de regras (T_1)*, ***w1thr*** = 0.1; *limiar mínimo para extração de regras (T_2)*, ***w2thr*** = 0.1.

Tabela 6.21 - Resultados da Evolving Fuzzy Neural Network – simulação 1

30 Partições	Erro de Treinamento	Erro de Teste	Erro de Classificação	Número de Regras
Média	0.3651	0.2195	0.0088	134
Desvio	0.0163	0.0462	0.0029	13.65

Como pode ser visto na Tabela 6.21, a rede Evolving Fuzzy Neural Network conseguiu uma boa generalização para o problema de classificação de padrões de odores. O erro de classificação médio obtido pelas 30 partições foi de 0.88%. Uma taxa de erro extremamente baixa, considerando os resultados das demais abordagens verificadas nesta dissertação. O menor erro de classificação encontrado foi de 0.51% na classificação dos cinco gases testados.

A Figura 6.1 apresenta uma projeção dos testes realizados nesta simulação. São projetadas simultaneamente as respostas fornecidas pela rede, para cada um dos gases, e suas respectivas saídas desejadas.

Como se pode verificar, o fato de não utilizarmos técnicas de poda ou agregação no treinamento da rede refletiu na grande quantidade de regras extraídas da rede nesta simulação. A média de regras de classificação extraídas da rede foi 134 regras, uma quantidade relativamente grande para apenas cinco classes. No próximo experimento será verificado se as técnicas de “abstração do conhecimento” sugeridas pelo modelo neuro-difuso diminuem a quantidade de regras extraídas da rede, mantendo a taxa de classificação estável.

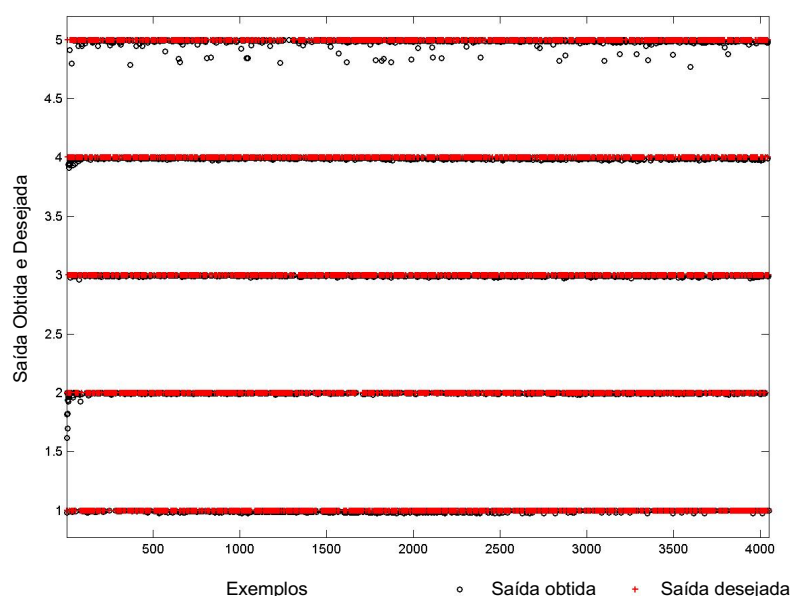


Figura 6.1 - Projeção das saídas obtidas e desejadas nos experimentos realizados com a Partição 1

No segundo experimento, apresentado na Tabela 6.22, a arquitetura Evolving Fuzzy Neural Network também foi inicializada sem nenhuma regra. Todos os nodos de regras foram criados dinamicamente durante o treinamento da rede. Porém, durante este treinamento é realizada a poda dos nodos de regras que estiverem fora do limiar de ativação estabelecido, ou que estejam dentro das regras de poda descritas na seção 4.2.2. A configuração dos parâmetros de treinamento foi realizada de maneira empírica e nos experimentos apresentados na Tabela 6.22, seus valores são os seguintes: $sthr = 0.6$; $errthr = 0.1$; $mf = 2$; $lr1 = 0.01$; $lr2 = 0.01$; $nodeage = 1000$; $aggintev = -1$; $maxfield = 0.5$; $w1thr = 0.1$; $w2thr = 0.1$.

Tabela 6.22 - Resultados da Evolving Fuzzy Neural Network– simulação 2

30 Partições	Erro de Treinamento	Erro de Teste	Erro de Classificação	Número de Regras
Média	0.5601	0.5082	0.0467	79
Desvio	0.0064	0.0643	0.0122	6.35

Com a aplicação da técnica de poda, a quantidade de nodos de regras na rede diminuiu e por consequência a quantidade de regras extraídas da rede neuro-difusa também decresceu. No experimento anterior a quantidade média de regras era de 134 regras, está média caiu para 79 regras na segunda simulação. Uma queda de 42% em relação ao número de regras geradas no primeiro experimento. O aumento médio no erro de classificação foi elevado, 531% superior em relação aos resultados da simulação anterior. Porém, este erro ainda é muito baixo, 4.67% sobre o conjunto de teste, o que não inviabiliza sua utilização.

A Figura 6.2 expõe a ação do mecanismo de poda durante o treinamento, pode-se verificar que durante a maior parte do treinamento uma grande quantidade de regras eram acrescentadas e posteriormente eliminadas da topologia da rede.

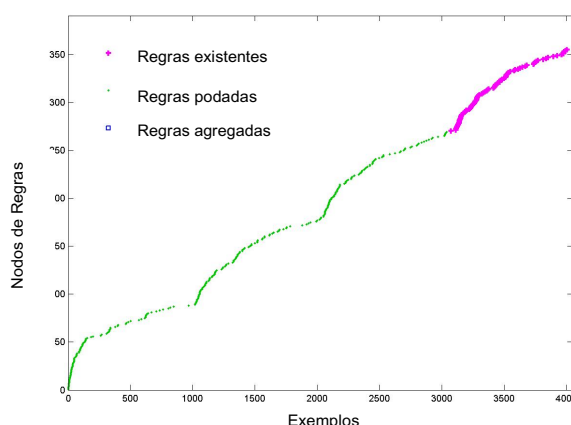


Figura 6.2 - Aplicação da técnica de poda nos experimentos com a Partição 2

A fim de verificar se o mecanismo de agregação poderia melhorar estes resultados (diminuir o erro de classificação em relação à segunda simulação e diminuir a quantidade de regras extraídas em relação à primeira simulação) e comprovar a capacidade da rede em aprender incrementalmente de maneira *on-line*. Neste experimento uma nova arquitetura Evolving Fuzzy Neural Network foi criada. Nesta nova arquitetura foram inseridas 10 regras, remanescentes de treinamentos prévios e novos padrões de odores foram apresentados a esta rede.

Este experimento nos permite verificar se o modelo realmente pode aprender ou se adaptar a novos padrões, sem perder ou “esquecer” o conhecimento previamente adquirido. Para isso, durante o treinamento da rede, foram utilizados os procedimentos de poda e agregação. Os resultados deste experimento são apresentados na Tabela 6.23. Os valores para cada um dos parâmetros da rede foram: $sthr = 0.9$; $errthr = 0.1$; $mf = 2$; $lr1 = 0.01$; $lr2 = 0.01$; $nodeage = 3000$; $aggintev = 100$; $maxfield = 0.5$; $w1thr = 0.1$; $w2thr = 0.1$.

Tabela 6.23 - Resultados da Evolving Fuzzy Neural Network– simulação 3

30 Partições	Erro de Treinamento	Erro de Teste	Erro de Classificação	Número de Regras
Média	0.4275	0.3717	0.0282	102
Desvio	0.0139	0.0551	0.0081	8.09

Nesta simulação houve uma melhora em relação aos resultados apresentados nos experimentos anteriores. O erro médio de classificação para as 30 partições testadas foi

de 2.82% e a quantidade média de regras foi de 102 regras. O menor erro de classificação obtido foi de 1,5%.

As figuras abaixo representam uma visualização da disposição dos padrões de odores e seus respectivos nodos de regras. As visualizações são feitas no espaço original dos dados, na Figura 6.3 e no espaço da Análise de Componentes Principais, na Figura 6.4. Nas duas figuras abaixo se pode perceber agrupamentos distintos de dados (Figura 6.4) e a correlação entre exemplos e nodos de regras (Figura 6.3).

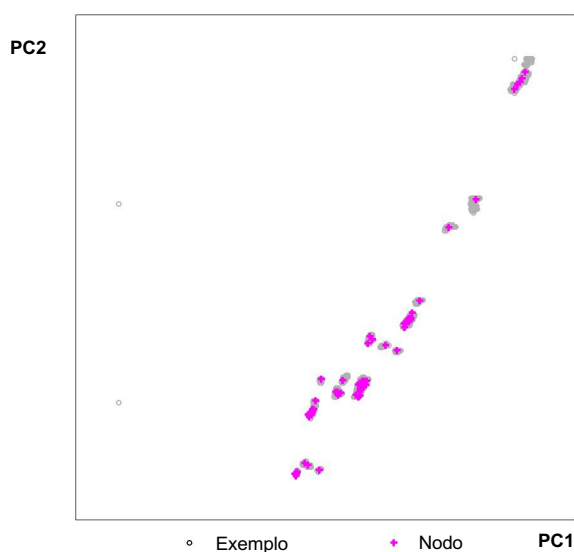


Figura 6.3 - Visualização dos padrões e seus respectivos nodos de regras, espaço original

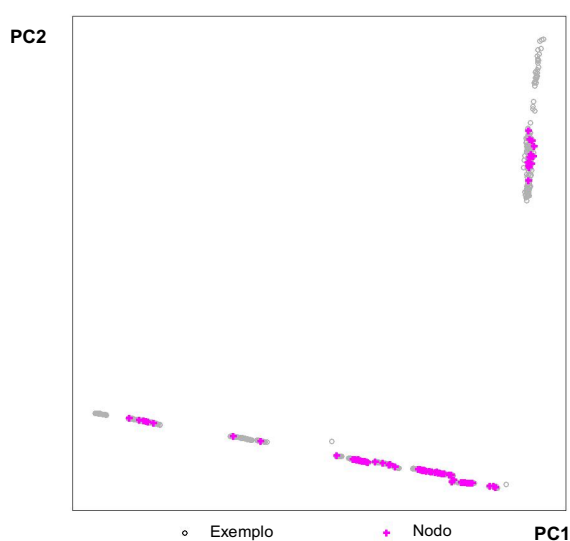


Figura 6.4 - Visualização dos padrões e seus respectivos nodos de regras, espaço PCA

Vale ressaltar que, este experimento possibilitou verificar além do treinamento *on-line*, a inserção de regras na topologia da rede. Pois, as regras utilizadas como conhecimento inicial na simulação, eram provenientes de um treinamento prévio e foram inseridas em uma arquitetura vazia de modo automático. Assim, o conhecimento pode ser inserido na rede Evolving Fuzzy Neural Network de duas formas distintas: a partir do treinamento da rede com novos exemplos ou com a inserção de regras previamente adquiridas.

O quarto experimento da série de simulações realizadas com a rede Evolving Fuzzy Neural Network visa verificar uma das mais interessantes de suas funcionalidades para o reconhecimento de odores: o treinamento *on-line* e a inserção dinâmica de classes na arquitetura da rede, sem a necessidade de treinar a rede com toda a base de dados novamente.

A simulação foi realizada da seguinte forma: a partir de uma arquitetura previamente treinada com quatro gases, escolhidos de forma aleatória, foram apresentados à rede exemplos de um quinto gás. Nenhum dos exemplos do quinto gás foi apresentado à rede na fase anterior de treinamento. A inclusão desta nova classe de exemplos no treinamento da rede, força o modelo a se adaptar a estes novos padrões e englobar o quinto gás automaticamente em sua topologia.

Para cada uma das partições, foram escolhidos diferentes gases a serem apresentados à rede na segunda etapa de treinamento. Assim, no primeiro experimento, a rede foi primeiramente treinada com os gases: Butano, Etano, Monóxido de Carbono e Propano. Na segunda etapa do treinamento (inserção do quinto gás) o gás Metano foi apresentado à rede.

No segundo experimento, primeiramente se treinou a rede com os gases: Etano, Monóxido de Carbono, Propano e Metano. O gás Butano foi apresentado na segunda fase do experimento. Para o terceiro experimento, os gases utilizados foram: Butano, Etano, Monóxido de Carbono e Metano na primeira fase do experimento. O gás Propano foi apresentado à rede posteriormente.

Os resultados destes experimentos são apresentados na Tabela 6.24 e os parâmetros utilizados no treinamento foram: $sthr = 0.9$; $errthr = 0.1$; $mf = 2$; $lr1 = 0.06$; $lr2 = 0.06$; $nodeage = -1$; $aggintev = -1$; $maxfield = 0.5$; $w1thr = 0.1$; $w2thr = 0.1$.

Tabela 6.24 - Resultados da Evolving Fuzzy Neural Network– simulação 4

Simulação	Erro de Treinamento	Erro de Teste	Erro de Classificação	Número de Regras
1	0.3785	0.3013	0.0126	161
2	0.4056	0.2378	0.0141	280
3	0.2469	0.1647	0.0066	228

Como pode ser verificado, o fato do quinto gás ter sido apresentado à rede posteriormente ao treinamento dos demais, não prejudicou em nada a generalização do modelo para os cinco gases. Os testes na rede neuro-difusa foram feitos com os cinco gases, simultaneamente. O erro médio de classificação desta simulação foi de 1.11%, porém a não utilização das técnicas de poda ou agregação pode ser percebida pela grande quantidade de regras extraídas da rede. A terceira simulação, apresentou o melhor resultado dos experimentos, um erro médio de classificação de 0.66%.

Abaixo é apresentada uma projeção das regras de classificação obtidas durante o treinamento da rede Evolving fuzzy Neural Networks com os padrões de odores. Estas regras foram obtidas no primeiro experimento.

Rule 1:

if [1] (1 0.524) (2 0.476)
 [2] (1 0.805) (2 0.195)
 [3] (1 0.788) (2 0.212)
 [4] (1 0.954)
 [5] (1 0.741) (2 0.259)
 [6] (1 0.851) (2 0.149)
 [7] (1 0.935)
 [8] (1 0.875) (2 0.125)
then (1 0.623) (2 0.372)

Rule 2:

if [1] (1 0.413) (2 0.587)
 [2] (1 0.759) (2 0.241)
 [3] (1 0.773) (2 0.227)
 [4] (1 0.955)
 [5] (1 0.743) (2 0.257)
 [6] (1 0.846) (2 0.154)
 [7] (1 0.891) (2 0.109)
 [8] (1 0.856) (2 0.144)
then (1 0.348) (2 0.651)

•
 •
 •
Rule 161:
if [1] (1 0.316) (2 0.684)
 [2] (1 0.615) (2 0.385)
 [3] (1 0.608) (2 0.392)
 [4] (1 0.954)
 [5] (1 0.591) (2 0.409)
 [6] (1 0.530) (2 0.470)
 [7] (1 0.853) (2 0.147)
 [8] (1 0.750) (2 0.250)
then (1 0.727) (2 0.273)

Considerando o que foi descrito na seção 4.2.2 do presente trabalho, os valores da premissa da regra entre colchetes [.], representam os atributos ou variáveis das instâncias. E os valores entre parênteses (.) representam respectivamente a função de pertinência (neste caso as variáveis lingüísticas utilizadas foram Baixo e Alto) e seu grau de pertinência associado para a referida variável. Da mesma forma para a conclusão da regra, a representação corresponde à função de pertinência e seu valor associado.

Desta forma, na regra 161, por exemplo, para o primeiro sensor o grau de pertinência associado à primeira função de pertinência é 0.316 e para a segunda função de pertinência é 0.684, da mesma forma para os demais atributos.

6.5 Comparação entre as abordagens

O objetivo desta seção é comparar o desempenho das abordagens investigadas na classificação dos padrões de odores. Nesta comparação são considerados os melhores resultados obtidos por cada um dos classificadores utilizados. Vale ressaltar que, para uma comparação idônea entre os classificadores, os experimentos foram realizados com a mesma base de dados. Para a comparação dos classificadores é utilizado o desempenho médio de classificação dos modelos nas partições testadas.

O parâmetro de comparação utilizado é o *Erro de Classificação* obtido por cada um dos classificadores. O Erro de Classificação é dado pela razão da quantidade de

padrões classificados de forma incorreta pela quantidade total de padrões testados. Para a realização do teste de hipóteses entre médias, são comparados os erros de classificação das 30 execuções realizadas com cada um dos classificadores.

A média dos melhores resultados obtidos por cada um dos classificadores, na classificação dos gases derivados do petróleo, são sumarizados na Tabela 6.25.

Tabela 6.25 – Melhores resultados obtidos pelos modelos investigados

Modelo	Partição	Erro de Treinamento		Erro de Validação		Erro de Teste		Erro de Classificação		Épocas Treinadas	
		Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
<i>MLP</i>	1	1749.77	2484.75	881.42	1235.83	898.56	1247.40	0.1388	0.1875	59	28
	2	2085.96	3292.44	1070.11	1685.21	1108.01	1656.09	0.1463	0.1742	71	41
	3	1914.03	3039.22	1039.65	1599.26	989.45	1574.15	0.1365	0.1726	73	36
<i>MLP Wavelet</i>	1	1249.58	2110.51	652.59	1064.61	626.92	1025.68	0.1029	0.1695	108	79
	2	1541.33	2468.88	810.58	1262.98	810.36	1272.35	0.1287	0.1938	74	25
	3	1486.09	2542.49	748.16	1299.52	756.73	1284.23	0.1135	0.1571	183	364
<i>TDNN</i>	1	3.22	2.88	75.61	28.83	63.99	15.94	0.0969	0.0881	18	21
	2	2.72	2.95	71.63	31.13	63.92	16.59	0.1261	0.0872	32	37
	3	2.81	2.92	72.90	30.60	64.14	16.38	0.1173	0.0865	27	32
<i>TDNN Wavelet</i>	1	1.34	1.98	27.17	21.87	66.49	6.99	0.0090	0.0383	193	229
	2	1.19	1.96	26.42	22.10	66.82	7.54	0.0068	0.0365	225	240
	3	1.23	1.97	26.36	22.05	66.93	7.63	0.0067	0.0365	207	236
<i>FWD</i>	1	134.35	5.78	-	-	570.99	36.92	0.2218	0.0615	5000	0
	2	177.61	0.07	-	-	430.00	0.0000	0.2000	0.0000	5000	0
	3	177.62	0.02	-	-	430.00	0.00	0.2000	0.0000	5000	0
<i>EFuNN</i>	<i>Média</i>	0.37	0.02	-	-	0.22	0.05	0.0088	0.0029	1	-

Na Tabela 6.26 é apresentada uma comparação estatística, baseada no teste de hipóteses *t* de *Student*, entre os classificadores investigados.

Tabela 6.26 - Comparação entre os classificadores

$H_0 : \mu_1 = \mu_2$			
$H_1 : \mu_1 < \mu_2$			
Modelo (μ_1)	Modelo (μ_2)	Valor calculado da variável de teste	Decisão
TDNN – Wavelet	MLP – Wavelet	-3.6146	Rejeitada
TDNN – Wavelet	FWD	-26.0622	Rejeitada
TDNN – Wavelet	EFuNN	-0.1903	Não Rejeitada
MLP – Wavelet	FWD	-3.1559	Rejeitada
MLP – Wavelet	EFuNN	3.6649	Rejeitada
EFuNN	FWD	-52.4949	Rejeitada

A Tabela 6.26 apresenta os resultados do teste t para amostras emparelhadas, com nível de significância de 5%. Nesta tabela, μ_1 e μ_2 são as médias de erros de classificação nos experimentos com cada um dos classificadores.

A seguir, o desempenho de cada uma das abordagens será discutida e seus resultados comparados com os demais modelos investigados.

6.5.1 Rede Perceptron Multi-Camadas

A rede Perceptron Multi-Camadas é um dos modelos conexionistas mais utilizados na classificação de padrões de odores. Parte por ser um dos modelos conexionistas mais difundidos e também por ser uma das abordagens mais simples. Sua utilização em narizes artificiais já foi amplamente investigada e seus resultados superaram os de muitas técnicas tradicionais na classificação de odores [Gardner et al., 1990; Gardner e Bartlett, 1994].

Na classificação dos gases, Butano, Etano, Metano, Propano e Monóxido de Carbono a rede Perceptron Multi-Camadas apresentou um erro médio de classificação de 11.5%, aliada ao pré-processamento com a Transformada de Wavelet.

Nos testes de hipótese realizados, apresentados na Tabela 6.26, pode ser verificado que a rede Perceptron Multi-Camadas só obteve um desempenho de classificação estatisticamente superior à rede Feature-weighted Detector. A

superioridade das demais arquiteturas sobre as redes Perceptron Multi-Camadas se deve a uma maior complexidade e robustez dos modelos (caso da rede EFuNN), além de uma melhor adaptabilidade aos dados que estão sendo tratados (caso da rede TDNN que possui a capacidade de aprender as características temporais dos padrões de odores).

6.5.2 Rede Neural com Atrasos no Tempo

A utilização das Redes Neurais com Atrasos no Tempo na classificação de padrões de odores foi proposta por Yamazaki (2001). Nos experimentos realizados pelo autor, esta abordagem apresentou melhoras significativas no reconhecimento de odores. Os testes foram realizados com amostras provenientes de três safras de um mesmo vinho tinto, perante os resultados apresentados pelas redes Perceptron Multi-Camadas na classificação da mesma base.

Nos experimentos realizados nesta dissertação, com os gases provenientes da indústria petrolífera, a Rede Neural com Atrasos no Tempo apresentou resultados similares ou superiores a rede Perceptron Multi-Camadas. Enquanto a rede Perceptron Multi-Camadas apresentou um erro médio de 11.5% a Rede Neural com Atrasos no Tempo apresentou um erro médio de 0.75%, utilizando pré-processamento com Análise Wavelet. Vale ressaltar que os objetivos deste trabalho não são investigar exaustivamente as técnicas abordadas e sim demonstrar suas potencialidades na classificação dos padrões de odores, fato pelo qual foi utilizado somente um atraso no tempo nos experimentos realizados com as Redes Neurais com Atrasos no Tempo.

Como pode ser verificado na Tabela 6.26, as Redes Neurais com Atrasos no Tempo correspondem a uma promissora abordagem para a classificação de odores, elas foram superiores em performance de generalização às redes Perceptron-Multi-Camadas e Feature-weighted Detector. As redes Evolving Fuzzy Neural Network apresentaram uma média de erro de 0.88%, estatisticamente similar as Redes Neurais com Atrasos no Tempo. Os excelentes resultados apresentados pela Rede Neural com Atrasos no Tempo são devidos a sua capacidade de utilizar as características temporais dos padrões de odores no aprendizado da rede.

6.5.3 Rede Feature-weighted Detector

As redes Feature-weighted Detector correspondem a uma nova abordagem híbrida, com o objetivo de resolver vários dos problemas encontrados no reconhecimento de padrões. Este modelo, introduzido por Li et al. (2001), foi utilizado como classificador de padrões de odor por se propor a classificar padrões, selecionar atributos e extrair regras explicativas da rede.

Apesar de ter apresentado bons resultados em experimentos anteriores [Amorin et al., 2003; Li et al., 2002], o modelo Feature-weighted Detector obteve estatisticamente o erro maior médio de classificação entre os classificadores investigados, um erro médio de 20.72%. A partir de maiores investigações com a arquitetura Feature-weighted Detector verificou-se que a mesma apresenta dificuldades para tratar dados não-linearmente separáveis.

Apesar da baixa performance de generalização da rede Feature-weighted Detector na classificação dos sinais de odor, a rede apresentou resultados promissores na identificação dos atributos relevantes para a classificação dos odores.

Verificar a importância de cada um dos sensores na classificação dos sinais de odores é uma tarefa muito importante. Mesmo se sabendo que o sistema olfativo biológico utiliza milhares de sensores para identificar a presença de um odor no ambiente; em seu correspondente eletrônico, o nariz artificial, quanto menor a quantidade de sensores utilizados, melhor será o conceito do dispositivo. Esta afirmação se deve pelo custo de aquisição de cada um dos sensores e pela necessidade do sistema de reconhecimento ser implementado em hardware, para o dispositivo ser portátil. No caso de uma rede neural, quanto menor a topologia da arquitetura a ser implementada, menor o custo desta tarefa. Outra questão a ser considerada é que a complexidade do reconhecimento de odores é diretamente proporcional a quantidade de sensores utilizados no reconhecimento do odor [Yamazaki, 2001].

Com a utilização do método de seleção de atributos da rede Feature-weighted Detector metade dos atributos da base foram removidos, sem prejuízo para a classificação dos padrões de odor. A rede Feature-weighted Detector identificou vários

dos atributos indicados como de maior relevância pelo método de Análise de Componentes Principais, porém realizando esta tarefa de forma automática e rápida.

Outras vantagens que pesam em favor da rede Feature-weighted Detector são sua simplicidade de implementação, a não necessidade de técnicas de otimização para as regras extraídas e o processo de extração de regras ser realizado de forma direta.

Sobre o processo de extração de regras difusas da rede, pode ser destacado que o mesmo é realizado de forma simples. Porém, somente podem ser extraídas uma regra de classificação por classe, o que pode prejudicar sua interpretação quando a quantidade de atributos da base de dados é muito grande. Outra deficiência do modelo Feature-weighted Detector é quanto ao método de seleção de atributos, apesar de ter apresentados bons resultados, este não possui uma metodologia bem definida, dando margem à subjetividade na escolha dos atributos menos relevantes.

6.5.4 Rede Evolving Fuzzy Neural Networks

Da mesma forma que a rede Feature-weighted Detector, são as funcionalidades da rede Evolving Fuzzy Neural Networks que a fazem se sobressair, perante abordagens conexionistas clássicas, na classificação dos sinais de odor. Estas funcionalidades são: (1) possibilidade de extração do conhecimento da rede a partir de regras difusas de classificação; (2) treinamento incremental e *on-line*, o que possibilita a apresentação de novos padrões e classes à rede, sem a necessidade de re-apresentar a base de dados toda no treinamento; e (3) extração, inserção e agregação do conhecimento da rede a partir da manipulação das regras difusas de classificação.

A média do erro de classificação obtido pela rede Evolving Fuzzy Neural Network é de 0.88%. Ao lado da Rede Neural com Atrasos no Tempo com pré-processamento Wavelet, são os melhores resultados na classificação de sinais de odores encontrados nos experimentos realizados nesta dissertação.

Além da boa performance de generalização, a rede Evolving Fuzzy Neural Networks apresenta algumas características ainda inexploradas na construção de narizes artificiais, como a aprendizagem incremental e *on-line*, as quais permitem que o

dispositivo aprenda novos odores durante a operação do nariz artificial. Esta flexibilidade faz com que o nariz artificial possa ser utilizado em diversas tarefas, sem que para isso o mesmo necessite ser reconfigurado ou alterado.

Outra característica da rede Evolving Fuzzy Neural Networks é a não necessidade de pré-processamento nos dados, esta funcionalidade pode não só diminuir o tempo de treinamento do dispositivo de reconhecimento de odores, como aumentar a velocidade na identificação dos gases detectados pelo sistema sensor. Outra vantagem da rede Evolving Fuzzy Neural Networks é que sua flexibilidade faz com que o sistema de reconhecimento de padrões se adapte facilmente a quantidade de sensores disponíveis para o reconhecimento do odor, tornando o dispositivo extremamente versátil.

Outro ponto a ser considerado sobre a rede Evolving Fuzzy Neural Networks é a possibilidade de inserção, extração e agregação de conhecimento da rede. A fácil manipulação das regras permite que o conhecimento adquirido por um nariz artificial possa ser repassado para outros dispositivos de forma extremamente simples. Esta característica possibilita que dois narizes artificiais trabalhem colaborativamente, a partir do compartilhamento de suas bases de conhecimento.

Uma das possíveis deficiências da rede Evolving Fuzzy Neural Network no tratamento de odores é o crescimento de sua arquitetura durante a operação do dispositivo. Se durante a tarefa de aprendizagem da rede Evolving Fuzzy Neural Network não for realizada poda ou agregação de unidade de regras, sua arquitetura pode tornar-se muito grande e complexa. Outra questão a ser considerada é que a rede Evolving Fuzzy Neural Networks possui um conceito de temporalidade não explícito, e sendo as informações temporais tão importantes no tratamento dos sinais de odor esta característica pode prejudicar seu desempenho.

6.6 Sumário do Capítulo

Este capítulo apresentou os resultados dos experimentos realizados com os classificadores investigados nesta dissertação. O objetivo de tais experimentos foi verificar qual das abordagens obtém melhor desempenho na classificação de estímulos odorantes captados por um nariz artificial.

A rede Perceptron Multi-Camadas obteve um erro médio de classificação nas três partições de 14.05%, porém, com a utilização do pré-processamento com a Transformada de Wavelets este erro caiu para 11.5%. Estes resultados foram obtidos testando-se diferentes topologias, utilizando uma camada intermediária e variando a quantidade de unidades de processamento nesta camada.

As Redes Neurais com Atrasos no Tempo obtiveram um erro de classificação de 11.34%. Com a utilização do pré-processamento Wavelet este erro caiu para 0.75%. O uso da técnica de Análise Wavelet, na redução do ruído e na compressão do sinal de odor, proporcionou as Redes Neurais com Atrasos no Tempo um dos melhores resultados na classificação dos sinais de odores.

As Redes Feature-weighted Detector obtiveram um erro médio de classificação alto em relação às demais abordagens, um erro de 20.72%. Porém, apresentaram bons resultados na seleção dos sensores mais importantes do nariz artificial, em sua análise quatro atributos foram identificados como de menor relevância na classificação dos sinais de odores. Sua utilização proporcionou uma diminuição de 50% do tamanho da base de dados e por consequência na complexidade do reconhecimento dos odores.

As redes evolutivas Evolving Fuzzy Neural Networks obtiveram ótimos resultados na classificação dos odores. O erro médio de classificação, nas três partições, foi de 0.88%. Além da boa generalização na classificação dos sinais de odores, a rede Evolving Fuzzy Neural Networks possui diversas funcionalidades que a tornam interessante na classificação de odores, como o aprendizado incremental e *on-line*, e a inserção, extração e agregação de conhecimento.

7

Conclusão

Este capítulo apresenta as considerações finais sobre o trabalho desenvolvido nesta dissertação. A seção 7.1 apresenta algumas considerações sobre os resultados obtidos nos experimentos. Na seção 7.2 são descritas as contribuições desta dissertação e a seção 7.3 apresenta algumas propostas para trabalhos futuros.

7.1 Considerações finais

A construção de dispositivos que simulem o comportamento do sistema olfativo biológico é uma área em expansão. Muitas pesquisas estão sendo despendidas com a finalidade de simular a percepção humana. A maioria dos esforços visa refletir nos dispositivos eletrônicos a capacidade humana de perceber o mundo na forma de cinco sentidos ou até mesmo estender nossas potencialidades sensoriais.

O presente trabalho está contextualizado na construção de narizes artificiais, dispositivos eletrônicos cuja necessidade em diversas aplicações é indiscutível. O foco principal desta dissertação foi à construção de sistemas de reconhecimento de padrões para narizes artificiais, buscando técnicas que apresentem resultados significativos na classificação de sinais de odor.

Foram apresentadas várias abordagens de sistemas de reconhecimento de padrões, como a rede Perceptron Múltipla-Camadas [Rumelhart et al., 1986], a rede Neural

com Atrasos no Tempo [Lang e Hilton, 1988] e algumas abordagens híbridas, como as redes Feature-weighted Detector [Li et al., 2002] e Evolving Fuzzy Neural Network [Kasabov, 1998b]. A base de odores utilizada no estudo de caso da presente dissertação é composta pelos gases derivados do petróleo Butano, Etano, Metano e Propano, também foi utilizado o gás Monóxido de Carbono nos experimentos.

Dentre as abordagens conexionistas, as redes Perceptron Multi-Camadas são as redes neurais artificiais mais utilizadas na literatura para classificação de sinais de odores [Yamazaki, 2001; Santos, 2000; Gardner e Bartlett, 1992; Gardner et al., 1994]. As Redes Neurais com Atrasos no Tempo, por sua vez, são o modelo conexionista que tem apresentado promissores resultados na classificação de odores [Zanchettin et al., 2003; Yamazaki et al., 2001]. Além das abordagens conexionistas, foi testada a utilização da Transformada de Wavelet [Daubechies, 1990] como método de redução do ruído inerente a captação do estímulo odorante e como método de compressão do sinal de odor.

Nos resultados obtidos com as arquiteturas conexionistas, a Rede Neural com Atrasos no Tempo confirmou sua superioridade em relação às redes Perceptron Multi-Camadas. Quando aliada ao método de pré-processamento com a Transformada de Wavelet a Rede Neural com Atrasos no Tempo obteve um erro médio de classificação de 0.75%, enquanto que a rede Perceptron Multi-Camadas obteve um erro médio de 11.5%.

Dentre as arquiteturas híbridas verificadas, a rede Feature-weighted Detector apresentou um desempenho de classificação baixo, em relação às demais abordagens investigadas na dissertação, estas redes apresentaram um erro de classificação médio de 20.72%. Porém, seu método de seleção de atributos demonstrou excelentes resultados, os quais podem ser comparados aos fornecidos pela técnica de Análise de Componentes Principais, na indicação dos atributos mais relevantes na classificação dos odores.

Com a utilização do método de seleção de atributos da rede Feature-weighted Detector foi possível identificar os sensores do nariz artificial que apresentam maior relevância na classificação dos gases testados. Também foi possível diminuir o tamanho da base de dados e conseqüentemente a complexidade da tarefa de classificação dos

odores, visto que esta é diretamente proporcional ao número de sensores utilizados na classificação dos odores.

Apesar do promissor resultado, na indicação dos atributos mais relevantes na classificação dos odores, o método de seleção de atributos da rede Feature-weighted Detector apresenta uma deficiência, a grande subjetividade na escolha dos atributos menos relevantes para a classificação. A rede Feature-weighted Detector também permite a extração de regras explicativas difusas da rede, porém somente pode ser extraída da rede uma regra por classe, o que prejudica a interpretação das regras em bases de dados com uma grande quantidade de atributos.

A rede Evolving Fuzzy Neural Networks por sua vez, apresentou resultados significativos na classificação dos padrões de odores, obteve um erro médio de classificação de 0.88%. Além da boa performance de generalização, existem outras características que tornam a rede Evolving Fuzzy Neural Networks preterida em relação as demais abordagens apresentadas, podem ser destacadas: a não necessidade de pré-processamento nos dados; seu método de aprendizagem incremental e *on-line*, que permite a rede aprender novos odores durante a operação do nariz artificial; a possibilidade de inserção, extração e agregação de conhecimento, os quais possibilitam que o conhecimento adquirido por um nariz artificial possa ser repassado a outros dispositivos de forma extremamente simples; e a extração de regras de classificação difusas, explicando o conhecimento adquirido pela rede durante o aprendizado.

A utilização de um sistema de reconhecimento de padrões como a rede Evolving Fuzzy Neural Networks, no reconhecimento de sinais de odor em um nariz artificial, permite a criação de equipamentos ao mesmo tempo versáteis e altamente específicos. O nariz artificial pode, por exemplo, ser configurado para identificar um odor específico e posteriormente treinado para identificar outros odores. Na maioria das abordagens para narizes artificiais somente características como erro de classificação são consideradas na avaliação dos modelos, as funcionalidades da rede Evolving Fuzzy Neural Networks sugerem uma ampliação nestes quesitos, buscando dispositivos cada vez mais versáteis.

Considerando métodos estatísticos de comparação, as melhores abordagens para a classificação de odores, testadas nesta dissertação, foram as Redes Neurais com Atrasos no Tempo e as redes Evolving Fuzzy Neural Networks.

Como para tarefas mais simples de reconhecimento já existe tecnologia suficiente para a detecção de odores, futuras pesquisas poderiam considerar outras funcionalidades na construção dos narizes artificiais, como a aquisição automática de conhecimento, configuração automática dos dispositivos, a praticidade e a versatilidade dos narizes artificiais.

7.2 Contribuições deste trabalho

Além dos resultados obtidos neste trabalho servirem de base para futuras comparações com novos sistemas propostos para reconhecer odores, outras contribuições podem ser destacadas como:

Utilização da Transformada de Wavelet no pré-processamento dos sinais de odor, como método de eliminação do ruído e compressão do sinal de odor.

Validação dos resultados previamente obtidos com as Redes Neurais com Atrasos no Tempo na classificação de sinais de odor, as mesmas somente haviam sido testadas no reconhecimento de odores com uma base de dados específica [Yamazaki, 2001].

Validação do desempenho da rede Feature-weighted Detector para uma nova base de dados, expondo suas qualidades e deficiências. A rede Feature-weighted Detector somente havia sido validada em duas aplicações [Li et al., 2002; Amorin et al., 2003].

Apresentação das potencialidades da rede Evolving Fuzzy Neural Networks na classificação dos sinais de odores, apresentando suas vantagens e funcionalidades aplicados ao funcionamento dos narizes artificiais.

7.3 Proposta para trabalhos futuros

Muitos tópicos, por motivos diversos, não foram incluídos neste trabalho. Por se encontrarem fora do escopo pretendido, por serem de interesse secundário, por serem muito extensos ou muito complexos para o objetivo da presente dissertação. Os principais desses tópicos são mencionados brevemente a seguir e podem ser estendidos em futuras investigações.

Considerando o desempenho apresentado pelas arquiteturas híbridas na classificação de odores, dentre as perspectivas futuras, está a possibilidade verificar o desempenho de outras abordagens híbridas na classificação dos sinais de odores, explorando o hibridismo em outros aspectos do problema.

Outra possibilidade seria explorar outros aspectos da Transformada de Wavelet na classificação dos odores, neste trabalho ela somente foi utilizada como método de compressão e diminuição do ruído inerente aos estímulos odorantes, captados pelo sistema sensor do nariz artificial.

Visto os promissores resultados e funcionalidades da rede Evolving Fuzzy Neural Network na classificação dos padrões de odor, uma das possibilidades é incorporar um método de seleção de atributos à rede. De forma que a rede possa identificar os sensores mais relevantes na classificação dos odores. Outra possibilidade de trabalho é utilizar as informações temporais obtidas na aquisição dos sinais de odor, no treinamento da rede Evolving Fuzzy Neural Network.

A partir da utilização da rede Evolving Fuzzy Neural Network como sistema de reconhecimento de padrões em narizes artificiais, outra interessante funcionalidade a ser adicionada a rede é a atualização automática dos parâmetros de configuração da rede, visto sua quantidade e dificuldade de otimização. Os parâmetros de configuração poderiam ser atualizados de forma automática a partir da evolução da rede, com técnicas como Algoritmos de Otimização Global [Glover e Laguna, 1997; Holland, 1975; Goldberg, 1989; Dowsland, 1993]. Esta automatização pode resultar em dispositivos de

reconhecimento de odores mais versáteis, com funcionamento independente e sem a necessidade de interferência humana.

Apêndice A

Análise de Componentes Principais

A Análise de Componentes Principais é essencialmente uma técnica de análise exploratória de dados, que visa à redução de sua dimensionalidade. A utilização da Análise de Componentes Principais objetiva indicar estruturas, passíveis de serem modeladas estatisticamente em processos similares àqueles estudados. Ou seja, a transformação de um conjunto original de variáveis em um conjunto menor de combinações lineares que representam a maior parte da variância do conjunto original dos dados.

Sendo assim, a Análise de Componentes Principais determina fatores (*componentes principais*) em seqüência, que explicam a maior parte possível da variância dos dados, com a menor quantidade possível destes fatores. Os componentes principais são extraídos de maneira que o primeiro componente principal contenha a maior quantidade da variação total dos dados.

De forma geral e bem simplificada, as combinações lineares das m -variáveis originais, que geram cada componente principal, podem ser representadas pela Equação (1).

$$PC_i = a_{i1}v_1 + a_{i2}v_2 + \dots + a_{im}v_m \quad (1)$$

Nesta equação, v_j (para $j = 1, 2, \dots, m$) são as variáveis originais e a_{ij} (para $j = 1, 2, \dots, m$) são os coeficientes que medem a importância de cada variável na i -ésima componente principal (CP_i). Ou seja, o *peso* que cada variável possui na combinação

linear. Os pesos nada mais são do que o co-seno do ângulo entre o eixo da componente principal e o eixo da variável original, desta forma seu valor estará sempre entre -1 e 1. Quanto mais próximo de 1 este valor estiver, maior é a influência que a variável possui na descrição desta componente principal.

Na Figura A.1 é apresentado um exemplo da extração dos componentes principais. O primeiro componente principal segue o sentido de maior variação dos dados, o segundo componente tenta expressar o restante da variação e assim sucessivamente. Teoricamente existem tantos componentes principais quanto for o número de variáveis nos dados.

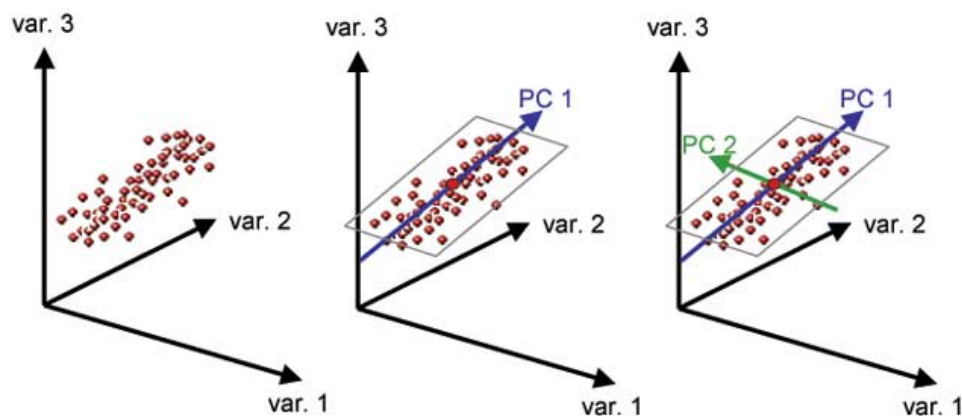


Figura A.1 - Exemplo da extração dos componentes principais [Daniels, 2003]

Definição do Método

Estas novas variáveis, chamadas componentes principais, são formadas de maneira que:

- ❑ As variáveis não são correlatas entre si (ortogonais);
- ❑ O primeiro componente principal explica o máximo da variação nos dados;
- ❑ Cada componente sucessivo explica o máximo possível da variação restante;

A análise de componentes principais pode ser resumida na seguinte equação:

$$L = V'RV \quad (2)$$

onde R é a matriz de correlação (poderia ser de covariância), V é a matriz de autovetores (*eigenvectors*) e L é a matriz de autovalores (*eigenvalues*).

Os autovalores são as raízes da equação polinomial definida por:

$$|\Sigma - \lambda \cdot I_n| = 0 \quad (3)$$

onde Σ é uma matriz de covariância ou de correlação, λ é o auto-valor e I é uma matriz identidade.

Os autovalores desta matriz obedecem as seguintes propriedades:

- O traço da matriz é igual a soma dos autovalores:

$$tr(\Sigma) = \sum_{i=1}^p \lambda_i \quad (4)$$

- O determinante da matriz é igual ao produto dos autovalores:

$$|\Sigma| = \prod_{i=1}^p \lambda_i = \lambda_1, \lambda_2, \dots, \lambda_p \quad (5)$$

A cada autovalor λ , os auto-vetores associados a λ são os vetores não nulos da solução do sistema:

$$(\Sigma - \lambda \cdot I_n)X = \bar{0} \quad (6)$$

onde Σ é uma matriz de covariância ou de correlação, λ é o autovalor, I é uma matriz identidade e X é o autovetor associado.

Os autovalores indicam a proporção da variação total explicada por cada combinação linear, enquanto que os autovetores fornecem os coeficientes das variáveis em cada função linear. Os componentes principais são perpendiculares entre si no espaço. Os valores dos componentes principais não são correlacionados entre si.

Para demonstrar a essência da Análise de Componentes Principais, será apresentado um exemplo simples que ilustra facilmente o método. Consideremos que cada amostra de um estudo contém uma concentração x de gás Hélio, uma concentração

y de gás Neônio, uma concentração z de gás Argônio, etc. Essencialmente, uma amostra desta base é caracterizada por um conjunto de n variáveis que correspondem a concentrações de gases nobres identificados. Se tomarmos o conjunto das amostras, formamos uma matriz de dados composta por n linhas (amostras ou gases) e m colunas (concentrações ou variáveis).

Para simplificar, imaginemos um conjunto de amostras caracterizadas por apenas duas variáveis (isto é, contendo concentrações pertencentes a apenas dois gases). Ao verificar a correlação entre elas, necessariamente é obtida uma reta de regressão linear que traduz a melhor correlação linear entre as duas variáveis. Se fosse possível definir uma nova variável que traduzisse fielmente a reta de regressão, esta variável poderia ser descrita como representante das duas primeiras variáveis. Ou seja, com esta nova variável poderíamos verificar o comportamento essencial das duas variáveis iniciais. Esta nova variável não seria mais do que um componente principal, que resulta da combinação linear entre as duas variáveis iniciais.

Este método estatístico de análise multivariada possui uma enorme utilidade quando se trabalha com amostragens caracterizadas por um número de variáveis bem superior a dois, como o apresentado nesta dissertação. Por exemplo, é muito útil partir de um conjunto de amostras com uma grande quantidade de variáveis e conseguir apenas algumas componentes que expliquem a maior parte da variância total dos dados. Como estas componentes podem descrever o comportamento do restante dos dados, elas podem ser utilizadas como parâmetros base para experimentos subjacentes com a base.

Apêndice B

Teste de Hipóteses

A distribuição de probabilidade das variáveis analisadas é determinada pela distribuição de probabilidade de cada uma das variáveis, cujas médias se quer comparar. A distribuição de probabilidade também é determinada pela escala numérica na qual tais variáveis estão expressas e pelo tipo de comparação que se deseja realizar. Assim sendo, existem diferentes tipos de testes que atendem diferentes tipos de variáveis, distribuições e comparações.

No que concerne à distribuição e natureza das variáveis a serem comparadas, existem basicamente dois grandes tipos de teste:

- ❑ Testes Paramétricos: São testes em que se pressupõe que as variáveis a comparar estão ambas, numa escala intervalar ou de razão, e que podem ser consideradas como apresentando um determinado tipo de distribuição de probabilidade (geralmente, a distribuição normal).
- ❑ Testes Não-Paramétricos: São testes em que se pressupõe relativamente pouca coisa acerca das variáveis envolvidas, baseando as comparações no ranking dos casos individuais.

Os testes paramétricos apresentam a vantagem de poder detectar diferenças muito sutis entre as variáveis estudadas, porém, sua aplicabilidade é limitada, devido à necessidade de se comprovar que todos os pressupostos do teste são de fato satisfeitos.

Já os não-paramétricos são comparativamente menos sensíveis, mas podem ser aplicados a um conjunto muito mais amplo de casos.

Tipos de Comparação

Existem dois tipos de comparação entre médias ou proporções que se pode realizar, cada um com suas particularidades e, conseqüentemente, exigindo um tipo específico de teste. Mais especificamente, tem-se:

- Teste Pareado: Comparação entre médias ou proporções de duas variáveis distintas;
- Teste Não-Pareado: Comparação entre médias ou proporções de dois subgrupos de casos de uma mesma variável.

Os testes pareados são aplicáveis quando se objetiva comparar duas observações de uma mesma amostra, sejam elas duas medidas separadas no tempo ou então duas dimensões diferentes do mesmo método. Já os testes não-pareados são usados quando se deseja comparar dois subconjuntos de uma amostra.

A Escolha de um Teste

A escolha de um teste de hipóteses depende diretamente dos objetivos da pesquisa, sendo válido, inclusive, tentar diversos testes até que um deles obtenha um achado estatisticamente relevante. Os testes de hipótese mais comuns são apresentados na Tabela B.1.

Tabela B.1 – Tipos de teste de hipótese

	Testes Paramétricos	Testes Não-Paramétricos
Testes Pareados	Teste <i>t</i> de Student Pareado	Teste <i>Wilcoxon</i> Teste <i>Signed Rank</i>
Testes Não-Pareados	Teste <i>t</i> de Student Não-Pareado	Teste de <i>Kolmogorov-SMironov</i> Teste <i>Mann-Whitney U</i>

Quando se objetiva detectar diferenças sutis, é melhor, quando possível, usar os testes paramétricos, já que esses são os mais sensíveis. Caso não seja obtido um

resultado significativo neles, é provável que os não-paramétricos também não o conseguirão. Em qualquer caso, porém, é preciso considerar o tipo de comparação a ser realizado, de modo que se possa escolher um teste pareado ou não-pareado conforme a indicação.

Conceitos básicos dos testes de hipótese

Hipótese nula e hipótese alternativa

Usualmente, devemos decidir entre duas hipóteses, uma sendo bastante específica a respeito do valor do parâmetro e a segunda fornecendo uma alternativa à primeira. A primeira hipótese costuma chamar-se hipótese nula e se denota por H_0 , a segunda é a hipótese alternativa, denotada por H_1 , conforme demonstrado na Equação (7).

$$\text{Hipóteses} \quad \begin{cases} H_0 : p = x \\ H_1 : p > x \end{cases} \quad (7)$$

onde p corresponde ao método que estamos verificando e x ao resultado fixado como desempenho de p antes do início do teste.

A hipótese nula, H_0 corresponde precisamente à suposição de que o novo método não tem desempenho melhor do que o método que está em uso, cujo desempenho é x . A hipótese alternativa, H_1 é $p > x$, porque se quer verificar se o desempenho do novo método é superior ao apresentado pelo método em uso.

Decididas as hipóteses à serem testadas, o próximo passo é construir um critério baseado no qual a hipótese nula será julgada. O critério de decisão é baseado na *estatística de teste* (que é uma estatística amostral, ou um valor baseado nos dados amostrais).

Região Crítica

É o conjunto de todos os valores da estatística de teste que levam à rejeição da hipótese nula.

Valor Crítico

É o valor, ou valores, que separam a região crítica dos valores da estatística de teste, estes valores não levam à região da hipótese nula.

Erros Tipo I e Tipo II

Há sempre um grau de incerteza na decisão tomada a respeito de uma hipótese estatística. Este é o preço a ser pago por se trabalhar em uma situação onde a variabilidade é inerente.

Ao se fazer um teste estatístico podemos cometer (não simultaneamente) dois tipos de erros.

- ❑ **Erro tipo I:** rejeitar H_0 , quando se deveria ter aceitado (α).
- ❑ **Erro tipo II:** aceitar H_0 , quando se deveria ter rejeitado (β).

Tabela B.2 – Tipos de erros

Conclusão do teste	Hipótese	
	H_0 é verdadeira	H_0 é falsa
Não rejeitar H_0	Correto	Erro tipo II
Rejeitar H_0	Erro tipo I	Correto

Para se decidir a favor ou contra H_0 , necessitamos de uma regra. Tal regra é chamada teste estatístico.

Testes bilaterais

As caudas em uma distribuição são as regiões extremas, delimitadas por valores críticos. Em um teste bilateral a região crítica (região de rejeição) está situada nas duas regiões extremas (caudas) sob a curva.

Nos testes bilaterais, o nível de significância alfa é dividido igualmente entre as duas caudas que constituem a região crítica. Por exemplo, em um teste bilateral, com nível de significância de 5%, há uma área de 0,025 em cada uma das caudas.

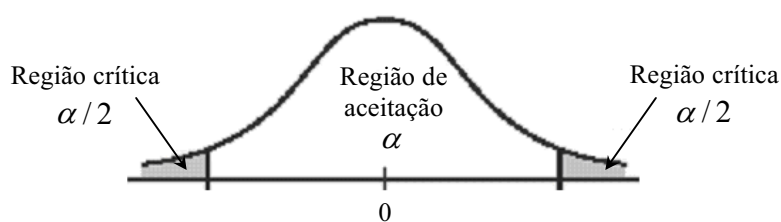


Figura B.1 – Teste de hipótese bilateral

No caso de comparação entre médias aritméticas, o objetivo do teste bilateral é verificar se duas amostras possuem diferenças significativas. Nesse caso, a hipótese nula é de que as médias aritméticas populacionais são iguais e a hipótese alternativa é que elas são diferentes, ou seja, $H_0: \mu_1 = \mu_2$ e $H_1: \mu_1 \neq \mu_2$.

Testes unilaterais

Os testes unilaterais avaliam situações do tipo "este método é melhor (ou pior) que aquele outro". A região de rejeição se localiza em uma única extremidade da curva da distribuição. Os testes unilaterais podem ser à esquerda ou à direita.

Testes unilaterais à esquerda

Os testes unilaterais esquerdos possuem a região crítica localizada na região extrema esquerda sob a curva. O objetivo do teste unilateral à esquerda é não só verificar se duas amostras possuem diferenças significativas, mas também indicar se a *amostra 1* apresenta resultados significativamente menores que a *amostra 2*. Nesse caso, a hipótese nula é que a média aritmética populacional da *amostra 1* é menor ou igual a da *amostra 2* e a hipótese alternativa é que a média aritmética populacional da *amostra 1* é maior que a da *amostra 2*, ou seja, $H_0: \mu_1 \leq \mu_2$ e $H_1: \mu_1 > \mu_2$.

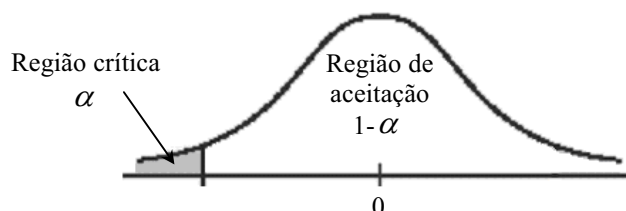


Figura B.2 – Teste de hipótese unilateral à esquerda

Testes unilaterais à direita

Os testes unilaterais direitos possuem região crítica localizada na região extrema direita sob a curva. O objetivo do teste unilateral à direita é não só verificar se duas amostras possuem diferenças significativas, mas também indicar se a amostra 1 apresenta resultados significativamente maiores do que a amostra 2. Nesse caso, a hipótese nula é que a média aritmética populacional da amostra 1 é maior ou igual a da amostra 2 e a hipótese alternativa é que a média aritmética populacional da amostra 1 é menor que a da amostra 2, ou seja, $H_0: \mu_1 \geq \mu_2$ e $H_1: \mu_1 < \mu_2$.

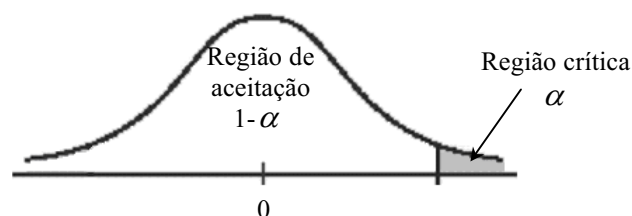


Figura B.3 – Teste de hipótese unilateral à direita

O Teste t de Student

O Teste t de Student é um teste de hipóteses paramétrico, sendo os seus pressupostos de que as duas variáveis envolvidas são intervalares e que ambas tem distribuição normal. Sendo tais pressupostos verdadeiros, a variável analisada resultante pode ser matematicamente transformada, de modo a formar a distribuição t de Student; em função da qual se determina se um dado resultado é ou não estatisticamente significativo.

O teste t de Student é um dos testes mais sensíveis que existem, podendo detectar diferenças muito pequenas, bastando apenas que se atenda aos seus requisitos. Além disso, existe tanto uma versão pareada quanto uma não-pareada.

O parâmetro usado para descrever a distribuição t é o número de graus de liberdade (g.l.), ou seja, o tamanho da amostra $(n) - 1$. Quando o tamanho da amostra cresce, a distribuição t se aproxima da normal, este comportamento é demonstrado na Figura B.4.

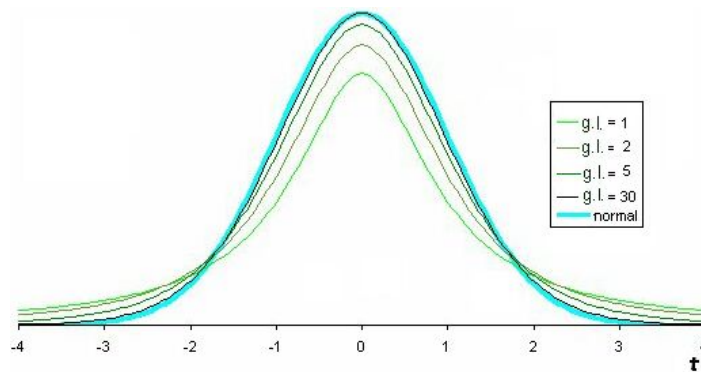


Figura B.4 – Distribuição t de Student

O teste é baseado na verificação de hipóteses, e por definição lógica, uma hipótese nunca pode ser afirmada, apenas refutada. Desta forma, inicialmente coloca-se a hipótese nula a ser confrontada. Na hipótese nula parte-se do princípio que não há efeito da variável independente sobre a dependente. No caso do teste de hipóteses para médias, parte-se do princípio de que não há diferença entre as médias de dois conjuntos de dados.

A seguir serão definidos os passos para realização do teste de hipóteses t de Student:

1. definir $H_0: \mu_1 = \mu_2$ e $H_1: \mu_1 \neq \mu_2$, $H_1: \mu_1 < \mu_2$ ou $H_1: \mu_1 > \mu_2$ dependendo do tipo de teste desejado. Onde, μ_1 e μ_2 são os valores médios reais das amostras A e B , respectivamente;

2. calcular os valores médios amostrais de A e B , denominados de $\bar{A}$ e $\bar{B}$;

3. calcular as variâncias amostrais de A e B , denominados de S_A^2 e S_B^2 ;

4. calcular $t = \frac{(\bar{A} - \bar{B}) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2 + S_B^2}{n}}}$ para testes bilaterais;

$$t = \frac{(\bar{A} - \bar{B}) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2 + S_B^2}{n}}} \geq t_{(1-\alpha), 2n-1} \text{ para testes unilaterais à esquerda;}$$

$$t = \frac{(\bar{A} - \bar{B}) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2 + S_B^2}{n}}} \leq t_{(1-\alpha), 2n-1} \text{ para testes unilaterais à direita;}$$

5. escolher um *nível de significância* (α), normalmente 1% ou 5%. O complemento do nível de significância ($1 - \alpha$) é conhecido como *nível de confiança* ou *intervalo de confiança*, e é normalmente de 99% ou 95%;
6. procurar na tabela *t-student* [Lehman, 1986] o valor $t_{(1-\alpha)/2, 2n-1}$ ou $t_{(1-\alpha), 2n-1}$, dado o *grau de liberdade* $2n-1$ e o nível de confiança $(1-\alpha)/2$ para o caso bilateral ou $(1-\alpha)$ para os casos unilaterais.
7. Caso os valores encontrados para t estejam entre o intervalo de $t > t_{(1-\alpha)/2, 2n-1}$ ou $t < -t_{(1-\alpha)/2, 2n-1}$ para o caso bilateral e $t > -t_{(1-\alpha), 2n-1}$ ou $t < t_{(1-\alpha), 2n-1}$ para os casos unilateral à esquerda e a direita, respectivamente, rejeita-se a hipótese nula. Em caso de rejeição da hipótese nula, os valores amostrais não condizem estatisticamente com o pressuposto na hipótese nula.

Referências Bibliográficas

- [Abraham e Nath, 2000] ABRAHAM, A. e NATH, B. (2000). **Hybrid Intelligent Systems Design: A Review of a Decade of Research**. technical report, School of Computing and Information Technology, Monash University, Australia, 39p. <http://citeseer.nj.nec.com/523093.html>. Acesso em 15 de ago. 2002.
- [Amorin et al., 2003] AMORIM, B. P.; ZANCHETTIN, C.; VASCONCELLOS, D. M. R. H.; VASCONCELOS, G. C.; LUDERMIR, T. B. e ARAÚJO, A. F. (2003). **Avaliação de um Modelo Neuro-Difuso para a Classificação de Padrões, Seleção de Atributos e Extração de Regras**. In: IV Encontro Nacional de Inteligência Artificial (ENIA), Brasil, p. 365-374, 2003.
- [Andrade, 2002] ANDRADE, A. F. (2002). **Redes Neurais**. Obtido em <<http://www.geoc.ufpr.br/~gfoto/pdi/neural/neural.html>>. Acesso em 31 de set. 2002.
- [Antonie et al., 1993] ANTOINE, J.; CARRETE, P.; MURENZI, R. e PIETTE, B. (1993). **Image Analysis with Two-dimensional Continuous Wavelet Transform**. In: Signal Processing, vol. 31, p. 241-272.
- [Barbosa et al., 2002] BARBOSA, M. S. S., LUDERMIR, T. B., SANTOS, F. L., SOUZA, J. E. G., SANTOS, M. S. e MELO, C. P. (2002). **Uma RBF para a Classificação de gases da Indústria Petrolífera**. In: VII Simpósio Brasileiro de Redes Neurais, Recife, Brasil, p.130-135.
- [Bishop, 1986] BISHOP, C.M. (1996). **Neural Networks for Pattern Recognition**. Oxford University Press, NY, 256p.
- [Braga et al., 2000] BRAGA, A. P., CARVALHO, A. P. L. e LUDERMIR, T.B. (2000). **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: Livros Técnicos e Científicos, 262p.

-
- [Brighan, 1988] BRIGHAM, E. (1988) **The Fast Fourier Transform and its Applications**. Englewood Cliffs: Prentice Hall, 448p.
- [Caruana et al., 2001] CARUANA, R.; LAWRENCE, S. e LEE GILES, C. (2001). **Overfitting in Neural Networks: Backpropagation, Conjugate Gradient, and Early Stopping**. Advances in Neural Information Processing Systems, Denver, Colorado. <<http://citeseer.nj.nec.com/caruana00overfitting.html>>. Acesso em 02 de set. 2003.
- [Daniels, 2003] DANIELS, R. (2003). **Electronic Noses**. Skin Care Forum, Issue 33 <http://www.scf-online.com/english/33_e/electronicnoses33_e.htm>. Acesso em 15 de mai. 2003.
- [Daubechies, 1988] DAUBECHIES, I. (1988). **Orthonormal Bases of Compactly Supported Wavelets**. In: Communications of Pure and Applied Mathematics, vol. 41, p. 909-996.
- [Daubechies, 1990] DAUBECHIES, I. (1990). **The Wavelet Transform, Time-frequency Localization and Signal Analysis**. In: IEEE Transactions on Information Theory, vol. 36, n. 5, p. 961-1005, September.
- [Daubechies, 1992] DAUBECHIES, I. (1992). **Ten Lectures on Wavelets**. In: CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia, Penssylvania, vol. 61, p. 909-996.
- [Di Natale et al., 1995] DI NATALE, C., DAVIDE, F. A. M., D'AMICO, A., SBERVEGLIERI, G., NELLI, P., FAGLIA, G. e PEREGO, C. (1995). **Complex Chemical Pattern Recognition with Sensor Array: the Discrimination of Vintage years of Wine**. In: Sensors and Actuators B, vol. 24-25, p. 801-804.
- [Donoho, 1995] DONOHO, D. L. (1995). **De-Noising by Soft-thresholding**. In: IEEE Trans. on Inf. Theory, vol. 41, n. 3, p. 613-627.
- [Dos Santos, 1999] SANTOS, F. L. (1999) **Caracterização Elétrica de Polímeros Condutores e de Sensores de Aroma**. Tese (Doutorado em Física), Curso de Pós-graduação em Física, Departamento de Física, Universidade Federal de Pernambuco, Recife-PE, 257p.
- [Dowsland, 1993] DOWSLAND, K.A. (1993). **Simulated Annealing**. In: Reeves, C.R. (ed), Modern Heuristic Techniques for Combinatorial Problems, Blackwell Scientific Publications, p. 20-69.
- [Duda et al., 2001] DUDA, R. O., HART, P. E., STORK e D. G. (2001). **Pattern Classification**. 2nd ed, Stork-John Wiley and Sons, 654p.

-
- [Elman, 1990] ELMAN, J. (1990). **Finding Structure in Time**. In: Cognitive Science, vol. 14, p. 179-211.
- [Gardner e Bartlett, 1992] GARDNER, J. W. e BARTLETT, P. N. (1992). **Sensors and Sensory Systems for an Electronic Nose**. NATO ASI Series E: Applied Sciences, Kluwer, Dordrecht, vol. 212, p. 346-351.
- [Gardner e Bartlett, 1994] GARDNER, J. W. e BARTLETT, P. N. (1994). **A Brief History of Electronic Noses**. In: Sensors and Actuators B, vol. 18-19, p. 211-220.
- [Gardner et al., 1990] GARDNER, J. W., HINES, E. L. e WILKINSON, M. (1990). **Application of Artificial Neural Networks to an Electronic Olfactory System**. In: Meas. Science Technology, vol. 1, p. 446-451.
- [Gardner et al., 1992] GARDNER, J. W., SHURMER, H. V. e TAN, T. T. (1992). **Application of an Electronic Nose to the Discrimination of Coffee**. In: Sensors and Actuators B, vol. 6, p. 71-75.
- [Gardner et al., 1994] GARDNER, J. W., PEARCE, T. C., FRIEL, S., BARTLETT, P. N. e BLAIR, N. (1994). **A Multisensor System for Beer Flavour Monitoring using an Array of Conducting Polymers and Predictive Classifiers**. In: Sensors and Actuators B, vol. 18-19, p. 240-243.
- [Gardner, 1991] GARDNER, J. W. (1991). **Detection of Vapours and Odours from a Multisensor Array Using Pattern Recognition**. Part 1: principal components and cluster analysis. In: Sensors and Actuators B, vol. 4, p. 108-116.
- [Glover e Laguna, 1997] GLOVER, F. and LAGUNA, M. (1997) **Tabu Search**. Kluwer academic Publishers, Boston, 408p.
- [Goldberg, 1989] GOLDBERG, D. E. (1989) **Genetic Algorithms in Search, Optimization and Machine Learning**. Massachusetts: Addison-Wesley Co, 432p.
- [Goonatilake e Khebbal, 1995] GOONATILAKE, S. e KHEBBAL, S. (1995) **Intelligent Hybrid Systems**. John Wiley & Sons, 340p.
- [Grossmann e Morlet, 1984] GROSSMANN, A. e MORLET, J. (1984). **Decomposition of Hardy Functions into Squared Integrable Wavelets of Constant Shape**. In: SIAM J. Math. Analysis, vol. 15, p. 723-736.
- [Haar, 1910] HAAR, A. (1910). **Theorie der Orthogonalen Funktionensysteme**. In: Mathematische Annalen, vol. 69, p. 331-371.

-
- [Hagan e Menhaj, 1994] HAGAN, M. T. e MENHAJ, M. B. (1994). **Training Feedforward Networks with the Marquardt Algorithm**. In: IEEE Transaction on Neural Networks, vol. 5, p.989-993.
- [Hall e Holmes, 1999] HALL, M. A. e HOLMES, G. (1999). **Benchmarking Attribute Selection Techniques for Data Mining**. Department of Computer Science, University of Waikato Hamilton, New Zealand, 15p.
- [Hayken, 1994] HAYKEN, S. (1994). **Neural Networks**, Prentice-Hall Inc., New Jersey et al, 842p.
- [Holland, 1975] HOLLAND, J. H. (1975). **Adaptation in Natural and Artificial Systems**. Ann Arbor, University of Michigan Press, 211p.
- [Jackson, 1991] JACKSON, J. E. (1991). **A User's Guide to Principal Components**. John Wiley and Sons, Inc., p. 1-25.
- [Jang, 1993] JANG, R. (1993). **ANFIS: Adaptive Network-Based Fuzzy Inference System**. In: IEEE Trans. on Syst., Man, Cybernetics, vol. 23, n. 3, May-June 1993, p. 665-685.
- [Kandel et al., 1998] KANDEL, A., ZHANG, Q.Y. e BURKE, H. (1998). **A Genetic Fuzzy Neural Network for Pattern Recognition**. In: IEEE Transaction on Neural Networks, Vol. 9, n. 1, p. 83-105.
- [Kasabov et al., 1997] KASABOV, N.; KIM J S; WATTS, M. e GRAY, A. (1997). **FuNN/2- A Fuzzy Neural Network Architecture for Adaptive Learning and Knowledge Acquisition**. In: Information Sciences - Applications, vol. 101, n. 3-4, p. 155-175.
- [Kasabov, 1988] KASABOV, N. (1988). **The ECOS Framework and the 'Eco' Training Method for Evolving Connectionist Systems**. In: Journal of Advanced Computational Intelligence, vol.2, n. 6, p. 1-8.
- [Kasabov, 1996] KASABOV, N. (1996). **Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering**. The MIT Press, Cambridge, MA, 550p.
- [Kasabov, 1996a] KASABOV, N. (1996). **Learning Fuzzy Rules and Approximate Reasoning in Fuzzy Neural Networks and Hybrid Systems**. In: Fuzzy Sets and Systems, vol. 82, n. 2, p. 2-20.

-
- [Kasabov, 1998] KASABOV, N. (1998). **ECOS: A Framework for Evolving Connectionist Systems and the Eco Learning Paradigm**. In: Proceedings of ICONIP'98, Kitakyushu, Japan, Oct98, IOSPress, p. 1232-1235.
- [Kasabov, 1998a] KASABOV, N. (1998). **Evolving Fuzzy Neural Networks - Algorithms, Applications and Biological Motivation**. In: Yamakawa T and Matsumoto G (Eds), Methodologies for the Conception, Design and Application of Soft Computing, World Scientific, p. 271-274.
- [Kasabov, 1998b] KASABOV, N. (1998). **Evolving Fuzzy Neural Networks - Algorithms, Applications and Biological Motivation**. In: Yamakawa T and Matsumoto G (Eds), Methodologies for the Conception, Design and Application of Soft Computing, World Scientific, p. 271-274.
- [Kasabov, 1998c] KASABOV, N. (1998). **Evolving Fuzzy Neural Networks - Algorithms, Applications and Biological Motivation**. In: Proceedings of Iizuka'98, Iizuka, Japan, p. 271-274.
- [Kasabov, 1999] KASABOV, N. (1999). **Evolving Connectionist and Fuzzy Connectionist System for On-line Decision Making and Control**. In: Soft Computing in Engineering Design and Manufacturing, Springer Verlag, 638p.
- [Kasabov, 1999a] KASABOV, N. (1999). **Evolving Connectionist and Fuzzy-Connectionist Systems: Theory and Applications for Adaptive, On-line Intelligent Systems**. In: N.Kasabov and R.Kozma (eds) Neuro-fuzzy Tools and Techniques for Intelligent Systems, Springer Verlag (Physica Verlag), p. 11-146.
- [Kasabov, 1999b] KASABOV, N. (1999). **Evolving Connectionist Systems for On-line, Knowledge-based Learning: Principles and Applications**. Technical Report TR99/02, Department of Information Science, University of Otago.
- [Kasabov, 2001] KASABOV, N. (2001). **On-line Learning, Reasoning, Rule Extraction and Aggregation in Locally Optimised Evolving Fuzzy Neural Networks**. In: Neurocomputing, Outubro 2001, vol. 41, p. 25-45.
- [Keller et al., 1994] KELLER, P. E., KOUZES, R. T. e KANGAS, L. J. (1994). **Three Neural Network Based Sensor Systems for Environmental Monitoring**. In: IEEE Electro 94 International Conference, Boston, MA, p. 378-382.
- [Keller et al., 1995] KELLER, P. E., KANGAS, L. J., LIDEN, L. H., HASHEM, S., e KOUZES R. T. (1995). **Electronic Noses and Their**

-
- Applications.** In: IEEE Technical Applications Conference and Workshop: Conference Record, p. 116-119.
- [Kress-Rogers, 1997] KRESS-ROGERS, E. (1997). **Biosensors and Electronic Noses for Practical Applications.** In: E. Kress-Rogers (Ed.), Handbook of Biosensors and Electronic Noses: Medicine, Food and the Environment, CRC Press, p. 3-39.
- [Lang e Hilton, 1988] LANG, K. J. e HILTON, G. E. (1988). **The Development of the Time-delay Neural Network Architecture for Speech Recognition.** Technical Report CMU-CS-88-152, Carnegie-Mellon University, Pittsburg, PA.
- [Lehman, 1986] LEHMAN, E. (1986) **Testing Statistical Hypothesis** (2nd ed.). Wiley, 600p.
- [Li e Hertz, 2000] LI, Z. e HERTZ, J. (2000). **Odour Recognition and Segmentation by a Model Olfactory Bulb and Cortex.** In: Computing Neural Systems, vol. 11, p. 83-102.
- [Li e Hopfield, 1989] LI, Z. e HOPFIELD, J. (1989). **A Model of the Olfactory Bulb and its Oscillatory Processing.** In: Biological Cybernetics, vol. 61, p. 379-392.
- [Li et al., 2002] LI, R. P., MUKAIDONO e M., TURKSEN, I. B. (2002). **A Fuzzy Neural Network for Pattern Classification and Feature Selection.** In: Fuzzy Sets and Systems vol. 130, p. 101-108.
- [Mallat, 1989] MALLAT, S. (1989). **A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.** In: IEEE Transactions on PAMI, n. 11, vol. 7, p. 674-693, July 1989.
- [Mallat, 1989] MALLAT, S. (1989). **A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.** In: IEEE Pattern Anal. and Machine Intell., v.11, n.7, p.674-693.
- [McCulloch e Pitts, 1943] MCCULLOCH, W.S. e PITTS, W. (1943) **A logical Calculus of the Idea Immanent in Nervous Activity.** In: Bulletin of Mathematical Biophysics, vol. 5, p. 115-133.
- [McGarry et al., 1999] MCGARRY, K.; WERMTER, S. e MACLNTYRE, J. (1999). **Hybrid Neural Systems: From Simple Coupling to Fully Integrated Neural Networks.** *Neural Computing Surveys*. <<http://citeseer.nj.nec.com/mcgarry99hybrid.html>>. Acesso em 12 de out. 2002.
- [Medsker, 1995] MEDSKER, L. R. (1995) **Hybrid Intelligent Systems.** Kluwer Academic Publishers, 298p.

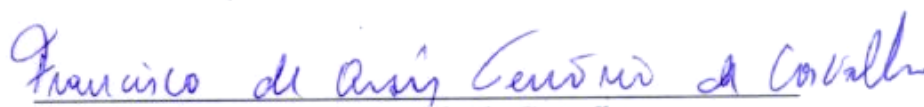
-
- [Misiti, 1996] MISITI, M.; MISITI, Y.; OPPENHEIM, G. e POGGI, J. M. (1996). **Wavelet Toolbox User's Guide for use with Matlab**. In: Natick, MA, The MathwWorks Inc, p.626.
- [Morlet, 1981] MORLET, J. (1981). **Sampling Theory and Wave Propagation**. In: Acoustic Signal/Image Processing and Recognition, ed. C. H. Chen (Springer-Verlag: Berlin), p. 233-261.
- [Nauck e Kurse, 1999] NAUCK, D. e KRUSE, R. (1999) **Neuro-fuzzy Systems for Function Approximation**. In: Fuzzy Sets and Systems, vol. 101, pp. 261-271.
- [Pekker e Janossy, 1986] PEKKER, S. e JANOSSY, A. (1986). **Chemistry of Doping and Distribution of Dopants in Polyacetylene**. In: Handbook of Conducting Polymers, vol. 1, ed. por T A Skotheim, Marcel Dekker, p. 45.
- [Persaud e Travers, 1997] PERSAUD, K. C. e TRAVERS, P. J. (1997). **Arrays of Broad Specificity Films for Sensing Volatile Chemicals**. In: E. Kress-Rogers (Ed.), Handbook of Biosensors and Electronic Noses: Medicine, Food and the Environment, CRC Press, p. 563-592.
- [Prechelt, 1994] PRECHELT, L. (1994). **Proben1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules**. Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe, Germany.
- [Prudêncio, 2002] PRUDÊNCIO, R. B. C. (2002). **Projeto Híbrido de Redes Neurais**, Dissertação (Mestrado em Ciência da Computação), Curso de Pós-graduação em Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife-PE, 101p.
- [Roitman, 2001] ROITMAN, V. L. (2001). **Um Modelo Computacional De Redes Neurais Para Redição Do Índice De Desemprego Aberto**. Tese (Doutorado em Ciências em Engenharia Civil), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 130p.
- [Rumelhart et al., 1986] RUMELHART, D. E., HINTON, G. E. e WILLIAMS, R. J. (1986). **Learning Representations by Backpropagation Errors**. In: Nature, vol. 323, p. 533-536.
- [Santos, 2000] SANTOS, M. S. (2000) **Construção de um Nariz Artificial usando Redes Neurais**, Tese (Doutorado em Ciência da Computação), Curso de Pós-graduação em Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife-PE, 120p.

-
- [Sundgren et al., 1991] SUNDGREN, H., WINQUIST, F., LUKKARI, I. e LUNDSTRÖM, I (1991). **Artificial Neural Networks and Gas Sensor Arrays: Quantification of Individual Components in a Gas Mixture**. In: Meas. Science Technology, vol. 2, p. 464-469.
- [Tabachnick e Fidell, 1996] TABACHNICK, B. e FIDELL, L. (1996). **Using Multivariate Statistics**. 3nd. ed., Ed. Harper Collins, 932p.
- [Vliet, 2003] VLIET, L. J. V. (2003). **Windowed Fourier Transform**. < www.ph.tn.tudelft.nl/~lucas/education/tn254/2002/Windowed%20Fourier%20Transform.pdf>. Acesso em 01 de set. 2003.
- [Waibel et al. (1989)] WAIBEL, A.; HANAZAWA, T.; HINTON, G.; SHIKANO, K. e LANG, K. (1989). **Phonemic Recognition using Time Delay Neural Networks**. In: IEEE Trans. Acoust., Speech, Signal Processing, vol. 37, n. 3, p. 328-339.
- [Yamazaki e Ludermir, 2001] YAMAZAKI, A. e LUDERMIR, T.B. (2001). **Classification of Vintages of Wine by an Artificial Nose with Neural Networks**. In: 8th International Conference on Neural Information Processing (ICONIP'2001), vol. 1, p. 184-187, Shanghai, China.
- [Yamazaki et al., 2001] YAMAZAKI, A., LUDERMIR, T.B. e DE SOUTO, M.C.P. (2001). **Classification of Vintages of Wine by an Artificial Nose using Time Delay Neural Networks**. In: IEE Electronics Letters, 22nd November, vol. 37, n. 24, p. 1466-1467.
- [Yamazaki, 2001] YAMAZAKI, A. (2001). **Reconhecimento de Padrões em um Nariz Artificial por Redes Neurais**, Dissertação (Mestrado em Ciência da Computação), Curso de Pós-graduação em Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife-PE, 113p.
- [Zanchettin e Ludermir, 2003a] ZANCHETTIN, C. e LUDERMIR, T. B. (2003). **Wavelet Filter for Noise Reduction and Signal Compression in an Artificial Nose**. In: Hybrid intelligent System, Melbourne, Austrália, 2003, p. 907-916.
- [Zanchettin e Ludermir, 2003b] ZANCHETTIN, C. e LUDERMIR, T. B. (2003). **A Neuro-Fuzzy Model Applied to Odor Recognition in an Artificial Nose**. In: Hybrid intelligent System, Melbourne, Austrália, 2003, p. 917-926.
- [Zanchettin e Ludermir, 2003c] ZANCHETTIN, C. e LUDERMIR, T. B. (2003). **Classification of Gases from the Petroliferous Industry by Artificial Nose using Evolving Fuzzy Neural Networks**. Relatório Técnico, 2003.

[Zanchettin et al., 2003]

ZANCHETTIN, C., LUDERMIR, T.B. e YAMAZAKI, A. (2003). **Classification of Gases from the Petroliferous Industry by an Artificial Nose with Neural Network**. In: Joint 13th International Conference on Artificial Neural Networks and 10th International Conference on Neural Information Processing (ICANN/ICONIP 2003), Istanbul, Turkey, p. 208-211.

Dissertação de Mestrado apresentada por **Cleber Zanchettin** a Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título , "**Sistema Neural Híbridos para Reconhecimento de Padrões em um Nariz Artificial**", orientada pela **Profa. Teresa Bernarda Ludermir** e aprovada pela Banca Examinadora formada pelos professores:


Prof. Francisco de Assis Tenório de Carvalho
Centro de Informática / UFPE


Prof. Adrião Duarte Dória Neto
Departamento de Engenharia Elétrica / UFRN


Profa. Teresa Bernarda Ludermir
Centro de Informática / UFPE

Visto e permitida a impressão.
Recife, 12 de março de 2004..


Prof. Jaelson Freire Brelaz de Castro
Coordenador da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federal de Pernambuco.